THE NATURAL HISTORY AND EPIDEMIOLOGY OF CERVICAL HUMAN PAPILLOMAVIRUS INFECTIONS IN MONTREAL UNIVERSITY STUDENTS

Harriet Richardson

Department of Epidemiology & Biostatistics
McGill University, Montreal

May 2003

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for a Doctor of Philosophy degree

ABSTRACT

Introduction. Only a small proportion of women with oncogenic HPV infections will eventually progress to high-grade squamous intraepithelial lesions (HSIL) or invasive cervical cancer (ICC), although the reasons why are not well understood. Additional knowledge about viral transmission and persistence is still needed, since some studies have shown that certain environmental co-factors, such as previous STDs or alcohol use may facilitate the sexual transmission of HPV or the persistence of an established HPV infection. The objectives of this study were to: 1) Describe the incidence and clearance rates of type-specific HPV infections; 2) Identify determinants of high- (HR) and low-risk (LR) HPV acquisition and clearance, and; 3) Identify viral determinants of low-grade squamous intraepithelial lesions (LSIL).

Methodology. In a prospective cohort of 621 Montreal university students, cervical specimens were collected for cytology and HPV DNA detection. Information on potential risk factors was obtained by interview at baseline, and at return visits. Follow-up visits were scheduled every 6 months over 2 years, for a total of 5 visits.

Results. The two-year cumulative incidence of any HPV infection was 36% and the mean duration of an episode with a type-specific HR- or LR-HPV infection was 16.3 and 13.4 months, respectively. After adjusting for age and sexual activity, co-factors for HPV acquisition included a recent *Chlamydia* infection, oral contraceptive use, alcohol use and washing after sex. Some determinants of HPV clearance included tobacco and alcohol use, in addition to use of tampons, daily vegetable consumption and condom use. Slightly different sets of the aforementioned co-factors or predictors were observed for HR- and LR-HPV infections. Non-European HPV16 or -18 variants appear to be strongly associated with incident low-grade squamous intraepithelial lesion (LSIL).

Conclusion. HPV infections occurred frequently in this cohort, and 24% or 12% of the women remained positive after 24 months with an incident type-specific HR- or LR-HPV infection. Some modifiable co-factors, independent of sexual activity, may facilitate transmission or persistence of certain HPV infections. These results may have implications for public health education and cervical cancer screening programmes.

RÉSUMÉ

Introduction. Seule une faible proportion des femmes infectées par des types oncogènes du virus du papillome humain (VPH) présentera une progression vers une lésion épidermoïde intra-épithéliale de grade élevé ou vers un cancer invasif du col utérin. L'explication de ce phénomène n'est pas encore connue. Il est nécessaire d'élargir les connaissances concernant la transmission et la persistance virale, puisque certaines études ont suggéré que certains co-facteurs environnementaux tels que des infections transmissibles sexuellement (ITS) ou la consommation d'alcool pourraient faciliter la transmission sexuelle du VPH ou la persistance d'une infection au VPH. Les objectifs de la présente étude étaient de : 1) décrire les taux d'incidence et d'élimination d'infections par le VPH; 2) identifier les déterminants de l'acquisition et de l'élimination d'infections par les types de VPH à risque cancérogène élevé et faible; 3) identifier les déterminants viraux pour les lésions épidermoïdes intra-épithéliales de bas grade (LSIL).

Méthodologie. Des échantillons de cellules du col utérin destinés à la cytologie et à la détection de l'ADN du VPH ont été prélevés chez cohorte prospective de 621 étudiantes montréalaises. L'information portant sur des facteurs de risque potentiels a été obtenue lors d'entrevues effectuées à la visite initiale et aux visites de suivi. Les visites de suivi étaient prévues tous les six mois durant deux ans, pour un total de cinq visites.

Résultats. L'incidence cumulative des infections par tous types de VPH pour la période de deux ans était de 36%. La durée moyenne d'un épisode d'infection par le même type de VPH était respectivement de 16,3 et 13,4 mois pour les VPH à risque cancérogène élevé et faible. Après ajustement pour l'age et l'activité sexuelle, les co-facteurs associés avec l'acquisition d'infection par le VPH étaient : une infection récente par Chlamydia, l'utilisation de contraceptifs oraux, la consommation d'alcool, ainsi que les ablutions après les rapports sexuels. Les déterminants de l'élimination d'infections par le VPH incluaient : la consommation du légumes, d'alcool et de tabac, l'utilisation de tampons et de condoms et la durée de la prise de contraceptifs oraux. Des profils de co-facteurs légèrement différents ont été observés pour les infections par VPH à risque cancérogène élevé et faible. Les infections par des variants non-européens de HPV-16 et -18 semblaient plus fortement associées avec les LSIL incidentes.

Conclusion. Les infections par VPH étaient communes dans cette cohorte; 24% des femmes ont conservé une infection incidente par un VPH à risque cancérogène élevé pendant 24 mois, alors que 12% des femmes ayant un VPH à risque faible l'ont conservé. Certains co-facteurs modifiables indépendents de l'activité sexuelle pourraient faciliter la transmission ou la persistance de certaines infections par VPH. Ces résultats sont pertinents pour l'éducation en santé publique ainsi que dans l'optique de programmes de dépistage du cancer du col utérin.

PREFACE

The format of this thesis follows that of a manuscript-based thesis. This dissertation consists of a collection of papers of which the student is an author or co-author. As per McGill University requirements, the papers must have a cohesive, unitary character making them a report of a single program of research. The structure for the manuscript-based thesis must conform to the following:

1. Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly duplicated text (not the reprints) of one or more published papers. These texts must be bound together as an integral part of the thesis. (Reprints of published papers can be included in the appendices at the end of the thesis.)

2. The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges between the different papers are mandatory.

3. The thesis must conform to all other requirements of the "Guidelines for Thesis Preparation" in addition to the manuscripts. The thesis must include the following: (a) a table of contents; (b) an abstract in English and French; (c) an introduction which clearly states the rationale and objectives of the research; (d) a comprehensive review of the literature (in addition to that covered in the introduction to each paper); (e) a final conclusion and summary.

4. As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g., in appendices) in sufficient detail to allow a clear and precise judgment to be made of the importance and originality of the research reported in the thesis.

5. When co-authored papers are included in a thesis the candidate must have made a substantial contribution to all papers included in the thesis. In addition, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. The supervisor must attest to the accuracy of this statement at the doctoral oral defense.

CONTRIBUTIONS OF CO-AUTHORS

Manuscript I

The natural history of type-specific human papillomavirus infections in female university students

Cancer Epidemiology, Biomarkers and Prevention 2003;12(6):485-90

Harriet Richardson, Ph.D Candidate:
Helped in the design of the study, coordinated the study and data collection, conceived the objectives of the analysis, designed and carried out statistical analyses and wrote the manuscript.

Gail Kelsall, study nurse/research assistant
Was responsible for subject recruitment, follow-up of participants, collection and transport of specimens and provided comments of the manuscript.

Pierre Tellier (M.D), Director, McGill University Student Health Services:
Granted access to the population attending the clinic and provided comments on the manuscript.

Hélène Voyer, Laboratory research assistant, Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal:
Was responsible for laboratory testing and provided comments on the manuscript

Michal Abrahamowicz, Researcher, Division of Clinical Epidemiology, Montreal General Hospital, and Associate Professor, Department of Epidemiology and Biostatistics, McGill University:
Contributed to the design of statistical analyses, provided statistical expertise and contributed to manuscript writing.

Alex Ferenczy (M.D), Professor, Department of Pathology, Jewish General Hospital, Faculty of Medicine, McGill University:
Co-designed the study, supervised the cervical cytology testing and provided comments on the manuscript

François Coutlée (M.D), Professor, Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal:
Co-designed study, coordinated and designed the laboratory testing procedures and contributed to manuscript writing

Eduardo L. Franco, Director, Division of Epidemiology, Department of Oncology, and Professor, Department of Epidemiology and Biostatistics, McGill University:
Designed and supervised the study, obtained funding for the investigation, participated in the design of statistical analyses and contributed to manuscript writing

Manuscript II

Co-factors for acquisition of low and high oncogenic risk cervical HPV infections in young women

(In preparation)

Harriet Richardson, Ph.D Candidate:
Helped in the design of the study, coordinated the study and data collection, conceived the object of the analysis, designed and carried out statistical analyses and wrote the manuscript.

Michal Abrahamowicz, Researcher, Division of Clinical Epidemiology, Montreal General Hospital, and Associate professor, Department of Epidemiology and Biostatistics, McGill University:
Contributed to the design of statistical analyses, provided statistical expertise and contributed to manuscript writing.

Pierre Tellier (M.D), Director, McGill University Student Health Services:
Granted access to the population attending the clinic and provided comments of the manuscript.

Gail Kelsall, study nurse/research assistant
Was responsible for subject recruitment, follow-up of participants, collection and transport of specimens and provided comments of the manuscript.

Roxane du Berger Research Associate, Division of Clinical Epidemiology, Montreal General Hospital,
Provided statistical expertise and contributed to manuscript writing.

Alex Ferenczy (M.D), Professor, Department of Pathology, Jewish General Hospital, Faculty of Medicine, McGill University:

Co-designed the study, supervised the cervical cytology testing and provided comments on the manuscript

François Coutlée (M.D), Professor, Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal:
Co-designed study, coordinated and designed the laboratory testing procedures and contributed to manuscript writing

Eduardo L. Franco, Director, Division of Epidemiology, Department of Oncology, and Professor, Department of Epidemiology and Biostatistics, McGill University:
Designed and supervised the study, obtained funding for the investigation, participated in the design of statistical analyses and contributed to manuscript writing

Manuscript III

Modifiable risk factors associated with clearance of type-specific cervical HPV infections

(In preparation)

Harriet Richardson, Ph.D Candidate:
Helped in the design of the study, coordinated the study and data collection, conceived the object of the analysis, designed and carried out statistical analyses and wrote the manuscript.

Michal Abrahamowicz, Researcher, Division of Clinical Epidemiology, Montreal General Hospital, and Associate professor, Department of Epidemiology and Biostatistics, McGill University:
Contributed to the design of statistical analyses, provided statistical expertise and contributed to manuscript writing.

Pierre Tellier (M.D), Director, McGill University Student Health Services:
Granted access to the population attending the clinic and provided comments of the manuscript.

Gail Kelsall, study nurse/research assistant
Was responsible for subject recruitment, follow-up of participants, collection and transport of specimens and provided comments of the manuscript.

Roxane du Berger Research Associate, Division of Clinical Epidemiology, Montreal General Hospital,
Provided statistical expertise and contributed to manuscript writing.

Alex Ferenczy (M.D), Professor, Department of Pathology, Jewish General Hospital, Faculty of Medicine, McGill University:

Co-designed the study, supervised the cervical cytology testing and provided comments on the manuscript

François Coutlée (M.D), Professor, Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal:
Co-designed study, coordinated and designed the laboratory testing procedures and contributed to manuscript writing

Eduardo L. Franco, Director, Division of Epidemiology, Department of Oncology, and Professor, Department of Epidemiology and Biostatistics, McGill University:
Designed and supervised the study, obtained funding for the investigation, participated in the design of statistical analyses and contributed to manuscript writing

Manuscript IV

Infection with HPV 16 or HPV 18 variants and risk of low-grade squamous intraepithelial lesions in young women

(In preparation)

Harriet Richardson, Ph.D Candidate:
Helped in the design of the study, coordinated the study and data collection, conceived the object of the analysis, designed and carried out statistical analyses and wrote the manuscript.

François Coutlée (M.D), Professor, Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal:
Co-designed study, coordinated and designed the laboratory testing procedures and contributed to manuscript writing

Michal Abrahamowicz, Researcher, Division of Clinical Epidemiology, Montreal General Hospital, and Associate professor, Department of Epidemiology and Biostatistics, McGill University:
Contributed to the design of statistical analyses, provided statistical expertise and contributed to manuscript writing.

Pierre Tellier (M.D), Director, McGill University Student Health Services:
Granted access to the population attending the clinic and provided comments of the manuscript.

Gail Kelsall, study nurse/research assistant
Was responsible for subject recruitment, follow-up of participants, collection and transport of specimens and provided comments of the manuscript.

Hélène Voyer, Laboratory research assistant, Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal:

Was responsible for laboratory testing and provided comments on the manuscript

Alex Ferenczy (M.D), Professor, Department of Pathology, Jewish General Hospital, Faculty of Medicine, McGill University:
Co-designed the study, supervised the cervical cytology testing and provided comments on the manuscript

Eduardo L. Franco, Director, Division of Epidemiology, Department of Oncology, and Professor, Department of Epidemiology and Biostatistics, McGill University:
Designed and supervised the study, obtained funding for the investigation, participated in the design of statistical analyses and contributed to manuscript writing

STATEMENT OF ORIGINALITY

This project was inspired by findings from an earlier cross-sectional study on risk factors for prevalent HPV infections. Results from that pilot study were part of my Master's thesis and the questions generated from that project helped to lay the groundwork for the development of a prospective cohort study at McGill and Concordia Universities. The latter was funded by a grant from the Canadian Institutes of Health Research to my research supervisor. I assisted him in the planning and preparation of that grant and helped with the design of the study instruments and informed consent and other patient forms.

There are very limited data on the incidence or persistence of HPV infections because of the time and cost involved in collecting data prospectively with multiple follow-up visits. At the initiation of this study there was only one published report that described the incidence of HPV infections and there was no knowledge about the average duration of HPV infections. Furthermore, only one study had focused on risk factors for HPV persistence at the time that this protocol was developed and it only evaluated persistence based on two visits.

In this study, the collection of data at multiple time points allowed for a complete description of incidence and duration of type-specific HPV infections over 2 years of follow-up. Moreover, to our knowledge, there is no published report that has evaluated potential risk factors for acquisition or clearance of different oncogenic groups of HPV, with an analysis that also accommodated changes in risk behaviour over time.

In this study DNA sequencing was used to identify variants of HPV-16 and HPV-18 to provide important and rare baseline data on the distribution of variants in this population. Knowledge about the frequency of these variants may prove useful in the future if certain HPV variants prove to be more pathogenic than others, as suggested in a handful of recent studies, and may need to be considered in screening or vaccination programs.

ACKNOWLEDGMENTS

Statement of Financial Support

This study was funded by a grant from the Canadian Institutes of Health Research (CIHR) (MT-13649).

I would like to sincerely thank the granting agencies for all the financial support I received over the course of my doctoral work and the opportunity it allowed me to pursue my academic goals. Over the course of my doctoral dissertation I was financially supported with a three year Doctoral Fellowship from the Canadian Institutes of Health Research and I received additional funding for two years from the CIHR study grant (MT-13649).

Dedication

This thesis would not have been possible without Dr. Eduardo Franco's unstinting support, encouragement and belief in my capabilities. I am very grateful for his willingness to entrust me with the responsibility of carrying out this study to completion and for the countless hours that he spent patiently reading and re-reading my work. I benefited enormously from his vast experience and substantive knowledge in the epidemiology of HPV infections and his passion for scientific research was both inspiring and motivating. Thank you.

Much of the success of the study is owed to the dedicated work of Gail Kelsall. Her commitment to the project and her warmth, charm and constant good spirits undoubtedly played an important role in earning the trust and confidence of the participants.

Dr. Pierre Tellier's committment to improving public health services and understanding the needs of the population he serves was a constant source of inspiration. His eagerness to collaborate with us greatly facilitated our access to the study population at McGill University and his charm and wit made it a most enjoyable experience.

I would also like to thank Suzanne Dumais and Louise Karline for their role in subject accrual at Concordia University and for ensuring high levels of compliance.

I wish to express my sincere thanks to Dr. Michal Abrahamowicz for his statistical expertise and guidance. It was both a privilege and a pleasure to have the opportunity to discuss the many nuances of my data with him. Our meetings were always lively, instructive and and most importantly, peppered with great humour! I would also like to thank his assistant Roxane du Berger for her help in some of the more complex analyses.

I must thank Dr. François Coutlée and his laboratory assistant Hélène Voyer for their meticulous and prompt processing of our samples. Their high standards and drive for perfection provided me with a great sense of confidence with regards to the HPV test results. I am also indebted to Dr. Coutlee's invaluable advice and feedback on many of the molecular biological aspects of this study.

I would like to thank many of my fellow colleagues and friends at the Division of Epidemiology, Department of Oncology for their contribution to a dynamic and intellectually stimulating work environment. In particular, I would like to thank Anita Koushik who thoroughly enriched my academic experience at McGill with many hours of stimulating discussions on the methodological issues of my study, and even more delightful hours of girl talk! I would also like to thank Susan Brogly for her valuable help in statistical programming and friendship. Candida Pizzolongo and Eliane Duarte Franco efficiently managed many of the administrative needs of this study and I would like to thank them for their significant contribution to the smooth functioning of our research team.

Words are inadequate to express my gratitude to my dear friend Marie-Claude Rousseau for her tremendous support and faith in my academic aspirations. Her thoughtful and intelligent feedback helped me to unravel many of the complexities of this study and I

will always be grateful for her willingness to graciously take time out of her own important pursuits to help a friend in need!

I would like to express my deepest thanks to my close friends, family and John who have patiently supported and encouraged me through this arduous process. I could not have done it without them. In particular, I would like to thank my mother for all of the editorial suggestions, the hours of help on my brutal tables and her unconditional love!

Finally, I would also like to thank Roche Molecular Systems who generously supplied us with all the reagents necessary for HPV genotyping.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 6: MANUSCRIPT II

CHAPTER 7: MANUSCRIPT III

CHAPTER 8: MANUSCRIPT IV

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background

Certain types of human papillomaviruses (HPV) are now believed to play a necessary role in cervical carcinogenesis [Schiffman et al., 1993; IARC, 1995; Bosch et al., 2002] and both longitudinal studies [Koutsky et al., 1992; de Sanjose et al., 1994; Ho et al., 1995; Nobbenhuis et al., 1999; Schlecht et al., 2001] and retrospective cohort studies with access to archival Pap smears [Wallin et al., 1999; Ylitalo et al., 2000] have confirmed that HPV precedes cervical precursor lesions and cervical cancer. There are approximately 40 HPV types that infect the genital tract and those HPV types most frequently found in cervical tumours are classified as high oncogenic risk (HR) HPV types. The remaining types are generally considered as low-oncogenic risk (LR) HPV types [Bauer et al., 1993; Bosch et al., 1995]. Variants for some HR-HPV types have been identified and classified based on their degreee of variation from the DNA sequence of the reference HR-HPV type.

Cervical infections with HPV are detected in 15%-40% of women with normal cervical cytology [Franco, 1991a; Herrero et al., 2000; Lazcano-Ponce et al., 2001]. However, when additional cervical specimens are taken from these women in follow-up surveys, the majority of the infections are found to be transient and only a small proportion of the women appear to harbour the same HPV type in subsequent specimens [Ho et al., 1995; Moscicki et al., 1998]. Thus, most cases of subclinical HPV infection appear to be transient and disappear naturally.

There have only been a few prospective cohort studies designed to investigate the dynamics of HPV acquisition or persistence in the last decade. Results from some of these studies show that the acquisition of new HPV infections can vary from 16 to 47 per 100 person-years [Van Doornum et al., 1994, Franco et al., 1999, Thomas et al., 2000]. The factors that most likely explain the different rates of HPV acquisition are the age distribution of each cohort and the sexual behaviour of the cohort members and their male partners [Ho et al., 1998; Kruger-Kjaer et al., 2001; Winer et al., 2003]. However,

while HPV infection appears to be primarily sexually transmitted, there is some evidence that other environmental co-factors such as tobacco and alcohol use and diet may play a role in facilitating HPV acquisition [Ho et al., 1998; Sedjo et al., 2002a; Sedjo et al., 2002b].

Furthermore, HR-HPV type may be associated with a different set of risk factors than LR-HPV types [Franco et al., 1995; Kjaer et al., 1997; Rousseau et al., 2000; Richardson et al., 2000; Chan et al., 2002], based on cross-sectional data. Therefore, certain associations may be diluted if HR- and LR-HPV infections are analyzed together. In addition, because prevalent infections are a mixture of persistent and newly acquired infections, there is still a need to have a better understanding of the co-factors that may facilitate the sexual transmission of a new HPV infection.

The median duration of an HPV infection appears to range from 8 months to 14 months in young and middle aged women [Ho et al., 1998; Franco et al., 1999; Woodman et al., 2001]. However, the majority of studies estimated the average duration of incident and prevalent infections combined. This strategy may result in an over-estimation of the average duration of a new HPV infection since prevalent infections are a mixture of established persistent infections and transient infections.

There is also substantial variation in the average duration, and the geographical distribution of specific HPV types. If the mean duration of an overall HPV infection depends strongly on the most prevalent HPV genotypes, estimates of duration may vary between studies because of differing frequencies of occurrence of HPV types in the respective cohort. Describing the geographical variation of genotypes and average duration of infection may help in establishing a clinically relevant definition of a persistent HPV infection that could have implications for cervical screening and HPV vaccination studies [Bosch et al., 2002].

The strongest predictors of a persistent HPV infection, thus far, include HPV type, viral load and the host immune response [Maciag et al., 2000]. However, very few studies

have studied the relationship between environmental co-factors and persistent HPV infection, and thus far, the effects of age, diet, tobacco and alcohol use, contraceptive practices and sexually transmitted diseases on HPV persistence are unclear and inconsistent [Hildesheim et al., 1994; Brisson et al., 1996; Moscicki et al., 1998; Ho et al., 1998]. However, the definition of HPV persistence has varied significantly between studies, and study designs have differed with respect to varying lengths of follow-up and number of return visits.

This project attempts to address some of these gaps in the literature. A cohort of 621 female students from McGill or Concordia University who were attending their respective University Health Clinic was established between November 1996 and November 1998. At enrollment, risk factor information was obtained by self-administered questionnaires that identified characteristics of sexual activity, hygiene practices, reproductive history, smoking and alcohol consumption, dietary habits and sociodemographic factors. In addition, a Pap smear and a cervical cell specimen were collected for HPV DNA testing by a consensus polymerase chain reaction (PCR) technique. Molecular variant analysis techniques were also used to sequence samples that were positive for HPV types 16 or 18 in an attempt to better characterize these specific infections. All subjects were followed for 2 years with 4 additional return visits scheduled every 6 months. At each visit repeat cervical specimens for cytology and HPV testing were collected and a modified questionnaire was administered to acquire information on aspects of recent sexual activity and other lifestyle behaviours. Thus, lifetime and recent predictors of HPV acquisition and persistence could be assessed using baseline and time-dependent covariates in multivariable regression models.

Because of the recent progress in diagnostic methods for detecting HPV DNA in cervical specimens many Western countries are now debating whether HPV testing should be used to augment current screening programs for cervical cancer, which are based on Pap cytology alone. Such an approach is sensible in principle, but because of the high prevalence of genital HPV infections, a positive HPV result carries a low positive predictive value with respect to cervical neoplasia, particularly for women under 30 years

of age [The ALTS Group, 2000]. In 1995, a position paper by the Canadian Task Force on Periodic Health Examination concluded that it was still premature to adopt large scale HPV testing in Canada [Johnson & Canadian Task Force on the Periodic Health Examination, 1995]. However, the Task Force did establish some research priorities as follows: (i) refinement of diagnostic methods, (ii) precise definition of HPV incidence in the population, (iii) assessment of risks associated with certain HPV genotypes for cancer progression, (iv) identification of cofactors influencing HPV transmission and carcinogenesis, (v) treatment of HPV infection, (vi) development of vaccines, and (vii) assessment of efficacy and cost-effectiveness of screening for HPV infection [Johnson & Canadian Task Force on the Periodic Health Examination, 1995]. This thesis addresses priorities ii-iv, outlined by the Task Force.

CHAPTER 2: REVIEW OF THE LITERATURE

2.1 The Evidence for a Relationship Between HPV and Cervical Cancer

2.1.1 Epidemiology of cervical cancer

Cervical cancer is the third most common cancer in women worldwide after breast and colorectal cancer, with a combined worldwide incidence of 371,000 new cases annually [Parkin et al., 1999]. The highest risk areas, with annual incidence rates greater than 40 per 100,000 women, are in Central and South America, Southern and Eastern Africa and the Caribbean [Ferlay et al., 1998]. In Canada, the yearly incidence rate for cervical cancer is approximately 8 new cases per 100,000 women with 1,500 new cases diagnosed annually, and 430 deaths per year attributed to the disease [Canadian Cancer Stats, 2000].

The past several decades have seen a decrease in morbidity and mortality due to cervical cancer in the industrialized nations. Improved perceptions about health and greater knowledge of associated risks have undoubtedly contributed to this observed decline, but it is the introduction of the Papanicolaou (Pap) test in the early 1950's [Papanicolaou, 1954] and the adoption of large-scale cytological screening that has contributed most significantly to the decline [Gustafsson et al., 1997]. Unfortunately, this decline has not been observed in most Third World countries where availability of Pap smear screening is limited to a small portion of the population (Franco, 1993).

Alarmingly, this trend is also observed in some sub-populations in more developed countries including Canada [Healey et al., 2001]. The highest rates of cervical cancer in Canada have been observed in the Northwest Territories (NWT), where cervical cancer is the most common malignant neoplasm among women and represented 35% of all cancers diagnosed between 1991 and 1996 [Corriveau, 1997; Gaudette et al., 1998]. Cancer rates are highest among the aboriginal residents who represent over 60% of the population in this region and the age standardized rate of cervical cancer in Native Aboriginal women in Saskatchewan is six times higher than the national rate [Franco et al., 2001]. Thus, the success of a Pap smear screening program appears to depend not only on a national commitment to support the necessary infrastructure to run a national cytological

screening program, but also depends on educating the public to overcome cultural barriers in order to avert a preventable disease.

2.1.2 Etiology of cervical cancer

As early as 1743 a French eighteenth century physician by the name of Jean Astruc hypothesized that uterine cancer (including cancer of the cervix) may be a sexually transmitted disease, postulating that a putative cause included "injection of semen tainted with lues" and "venereal virus" [Skrabanek, 1988]. Two hundred years later, Gagnon observed that cervical cancer was extremely rare in virgins and nuns [Gagnon, 1950]. Other early studies looking at environmental factors of cervical cancer, observed that women with cervical cancer had lower socio-economic status, were married earlier and had their first sexual encounter at an earlier age than controls [Wynder et al., 1954; Jones et al., 1958; Boyd & Doll, 1964]. Since then, the association of cervical cancer and sexual activity has been the most consistent epidemiological finding, irrespective of study design and methods of analysis [Bosch et al., 2002]. Furthermore, case-control studies examining the role of the 'male factor' have identified a significant increase in the risk of cervical cancer with the number of sexual partners reported by the male partner/husband. In all of these studies, the husbands of cases reported significantly more sexual partners than husbands of controls [Brinton et al., 1989; Bosch et al., 1994; Bosch et al., 1995; Juarez-Figueroa et al., 2001; Thomas et al., 2001].

2.1.3 HPV and cervical carcinogenesis

The consistently strong association between cervical cancer and sexual behaviour, in particular, number of lifetime sexual partners and age at first sexual encounter, [Rotkin, 1967; Martin, 1967; Pridan & Lilienfeld, 1971; La Vecchia et al., 1986; Brinton et al., 1987; Brisson et al., 1988; Brinton, 1992; Brisson et al., 1994] prompted the suggestion that a sexually transmitted, infectious agent may have an etiological role in cervical carcinogenesis [Franco, 1991b]. In 1976, zur Hausen observed that cervical cancer and papillomavirus induced condyloma acuminata (genital warts), shared similar epidemiological profiles, and postulated that a papillomavirus may be involved in the development of cervical cancer [zur Hausen, 1976]. The last two decades have since seen

an explosion of epidemiological and experimental research, all implicating human papillomavirus (HPV) as the most likely cause of cervical cancer. The biological evidence is summarized in table 2.1.

Table 2.1
Biological evidence for a carcinogenic role of HPV

- Papillomaviruses associated with tumours in other species
  Campo et al. (1980)
- Transcriptionally active viral genome frequently found in cervical tumors
  Schwarz et al. (1985)
- HPV found in both early and advanced tumour lesions
  Schneider et al. (1987a)
- Malignant transformation upon transfection of viral DNA into cultured cells
  Bedell et al. (1987)
- Cooperation between HPV and oncogenes in transformation
  Matlashewski et al. (1987)
- Viral DNA integrated in malignant lesions and episomal in premalignant lesions
  McCance et al. (1988)
- HPV E7 protein binds to the retinoblastoma anti-oncogene product
  Dyson et al. (1989)
- HPV E6 binds to the p53 tumor-suppressor gene product
  Werness et al. (1990)
- Cervical cells immortalized *in vitro* with HPV DNA differentiate dysplastically *in vivo* after being implanted in nude mice
  Waggoner et al. (1990)
- HPV DNA identified in 99.7% of ~1000 cervical cancer specimens in an international collaborative study
  Walboomers et al. (1999)

In brief, papillomaviruses were already known to cause tumours in a variety of mammalian species [Campo et al., 1980]. HPV was discovered in both early and advanced cervical tumour lesions [Schneider et al., 1987]. Malignant transformation was demonstrated upon transfection of viral DNA into cultured cells *[Bedell et al., 1987]* and found to depend on HPV E6 and E7 gene expression [von Knebel-Doeberitz et al., 1988; Crook et al., 1989]. The proteins encoded by the E7, E6 and E5 domains of the HPV genome were shown to bind to the products of three very important cell growth regulators including the retinoblastoma [Munger et al., 1989; Dyson et al., 1989] and p53 tumour-suppressor genes, [Werness et al., 1990; Vousden, 1993] and the epithelial growth factor (EGF) receptor [Banks & Matlashewski, 1993], respectively, resulting in loss of control at essential points in the cell cycle and amplification of certain cell signal transductions.

A panel of expert scientists convened by the World Health Organization's International Agency for Research on Cancer (IARC) concluded in 1995 that there was enough compelling evidence, both from biological and epidemiological data, to classify HPV-16 and HPV-18 as human carcinogens and HPV 31 and HPV 33 as probable carcinogens [IARC, 1995]. Other HPV types were classified as possibly carcinogenic with the exception of HPV types 6 and 11.

A variety of case-control and cohort studies conducted in the last decade have consistently shown that HPV infection is the strongest risk factor for cervical cancer, with relative risks (RR) in the range of 20-70 [Bosch et al., 2002]. Recently, the results of the combined analyses of the IARC based multicentric case control study have been described [Munoz et al., 2000] and the combined odds ratio (OR) for HPV detection was 83.3 (95% CI 54.9, 105.3), based on 2288 cases and 2513 controls from seven countries and using PCR for HPV detection and typing. With improved methods for viral nucleic acid detection, more than 90% of human cervical carcinomas, including condylomata, intraepithelial neoplasms and invasive tumours were found to harbour HPV DNA sequences compared to 5-20% of specimens from controls [Munoz et al., 1992] and in a more recent international collaborative effort, HPV DNA was found to be present in 99.7% of cervical cancer specimens [Giannoudis & Herrington, 2001]. Longitudinal

studies and retrospective cohort studies with access to archival Pap smears have confirmed that HPV infection precedes both cervical precursor lesions and cervical cancer [Koutsky et al., 1992; de Sanjose et al., 1994; Ho et al., 1995; Wallin et al., 1999; Nobbenhuis et al., 1999; Ylitalo et al., 2000; Schlecht et al., 2001].

More recently, human papillomavirus (HPV) DNA has been measured in exfoliated cells from the penile shaft, the coronal sulcus and the distal urethra, in studies investigating the role of the male partner in HPV transmission [Kjaer et al., 1991; Bergman & Nalick, 1992; Bosch et al., 1996; Castellsague et al., 1997]. Poor hygiene of the male partner has also been thought to play a role in the etiology of cervical cancer, with special attention given to the effects of circumcision. Recent investigations have shown that circumcized men were less likely to be HPV carriers [Castellsague et al., 2002b] and that wives of circumcised men had a significantly lower risk of cervical cancer [Kjaer et al., 1991]; results that correlate with the low incidence rates of cervical cancer observed in populations that practice circumcision such as found in the Middle East [Pridan & Lilienfeld, 1971].

2.1.4 Environmental co-factors for cervical carcinogenesis

Other risk factors have been identified for cervical precursor lesions, such as low-grade squamous intraepithelial lesions (LSIL), high-grade squamous intraepithelial lesions (HSIL) or invasive cervical cancer, independent of sexual activity or among HPV-positive subjects, although the results from epidemiological studies have not been consistent. Hormonal influences are likely to play a role in HPV carcinogenesis, and both increased parity and duration of oral contraceptive use (> 5 years) [Lacey et al., 1999; Moreno et al., 2002; Munoz et al., 2002] have been associated with a four-fold increased risk for cervical cancer.

Similarly, increased tobacco consumption has been related to cervical cancer [Castellsague et al., 2002b] and in studies that restricted the analyses to HPV positive women only, current smokers were about 2-3 times more likely to have cervical cancer or earlier precursor lesions compared to women who never smoked [Szarewski & Cuzick,

1998; Castellsague et al., 2002b]. Results from preliminary experimental evidence suggests that some high-risk (HR-HPV) infected cells may be more susceptible to DNA damage from specific tobacco carcinogens such as benzo[a]pyrene (B[a]P) [Melikian et al., 1999]. However, other authors argue that exposure to tobacco may affect the host immunosurveillance against viral infections, instead, as it has been shown that smoking may reduce the number of Langerhans cells and other markers of immune function [Poppe et al., 1995; Castellsague et al., 2002b].

Furthermore, conditions leading to immunosuppression such as an HIV infection can greatly increase the risk of cervical cancer and in 1993, the Centres for Disease Control (CDC) classified cervical cancer as an AIDS defining illness in women infected with HIV. Other sexually transmitted diseases such as Herpes simplex virus 2 (HSV-2) and *Chlamydia trachomatis* have inconsistently been associated with cervical cancer. These associations may be biased because of residual confounding due to similar risk profiles to that of HPV. However, other STDs may also reflect a lowered immune response or they may modify the carcinogenic potential of HPV via a biological interaction or stimulation of an inflammatory response [Anttila et al., 2001; Smith et al., 2002].

2.1.5 Classification of cervical precancerous lesions

Cytological screening for the detection of premalignant lesions and/or cervical cancer was first introduced by Papanicolaou in 1954 [Papanicolaou, 1954]. Commonly referred to as the Pap smear test, it is based on the recognition of cytomorphological signs indicating disruption in the maturation of the cervical squamous epithelium [Koss, 1989]. The grading system of cervical lesions was based on the concept of cervical intraepithelial neoplasia (CIN) and dysplasia [Richart, 1968]. Cervical lesions were classified into three grades of CIN according to the severity of disturbances of cellular maturation, stratification and cytological atypia. CIN I was defined as a well-differentiated intraepithelial neoplasm, now often associated with HPV 6 and 11, while CIN II and III were defined as poorly differentiated intraepithelial neoplasms, most often associated with HPV-16 and -18. In 1989, the CIN classification was replaced by the Bethesda classification system. A new atypia category was defined to describe atypical

squamous cells of undetermined significance (ASCUS) and the three CIN grades were re-classified as low-grade and high-grade squamous intraepithelial lesions (LSIL and HSIL), as described in table 2.2. Low-grade SIL represents CIN grade I corresponding to those lesions with a low risk of progression to carcinoma and high-grade SIL includes both CIN grade II and III corresponding to lesions with increased likelihood of progressing to carcinoma [Solomon, 1989].

Table 2.2
Bethesda classification of cellular abnormalities of the cervix

Designation	Abnormalities
ASCUS (atypia)	Atypical squamous cells of undetermined significance (ASCUS) Excludes preneoplastic changes
Low-Grade SIL* (LSIL)	1. Cellular changes associated with HPV (without other abnormalities) 2. Mild dysplasia (originally termed CIN I**)
High-Grade SIL (HSIL)	1. Moderate dysplasia (originally termed CIN II) 2. Severe dysplasia (originally termed CIN III) 3. Carcinoma *in situ* (originally termed CIN III) 4. One of the latter abnormalities and HPV associated changes

**SIL=Squamous intraepithelial lesion, **CIN=Cervical intraepithelial neoplasia*

2.1.6 Occurrence of cervical lesions

Follow-up studies of women with and without cervical abnormalities at enrollment have indicated that the continuous presence of a HR-HPV infection, especially HPV 16 or -18, is necessary for the maintenance and progression of CIN disease [Koutsky et al., 1992; de Sanjose et al., 1994; Ho et al., 1995; Bosch et al., 2002; Nobbenhuis et al., 1999; Schlecht et al., 2001; Moscicki et al., 2001]. Women identified with ASCUS or CIN I but HPV negative are likely to regress to normal instead of progressing to CIN II/III [White et al., 1998; Nobbenhuis et al., 2001]. In contrast, approximately 15-30% of cytological normal women with a HR-HPV infection have been observed to develop CIN II or III within 4

years of follow-up [van Staveren et al., 1986; Koutsky et al., 1992; Rozendaal et al., 1996]. However, the 3 year cumulative incidence of 0.15 (95% CI: 0.13-0.17) for LSILs among HPV positive women (13-20 years) attending family planning clinics in California [Moscicki et al., 2001] was much lower than the 2 year cumulative incidence of 0.28 (95% CI: 0.10-0.47) for an LSIL, from time of first HPV positive test, among an older cohort of women between 16-50 years of age, recruited from STD clinics in Seattle [Koutsky et al., 1992]. Potential differences in the proportion of HPV persistors and/or the distribution of age and sexual behaviour between these two cohorts could explain the higher risk observed in the Seattle cohort.

2.1.7 Screening for cervical cancer

Cytology screening has been shown to be very effective in reducing the incidence and mortality of cervical cancer in Canada [Storey et al., 1998], Scandinavia, Western Europe and the USA [Gustafsson et al., 1997]. However, a meta-analysis of a large number of studies indicated that the sensitivity of a single Pap smear is only 51% while the specificity is 98% for detection of a high-grade lesion [McCrory et al., 1999]. About one-third of the false negative diagnoses are attributable to slide interpretation errors and another two-thirds to poor sample collection and slide preparation [McCrory et al., 1999]. The current success of the Pap smear is based on the recommendation that women have a cervical cytological examination annually. With every additional Pap smear a woman has, the probability of detecting a cervical lesion increases, provided that a cervical lesion is present [McCrory et al., 1999].

2.1.8 HPV testing

HPV testing has also been suggested as a tool for improving the sensitivity and specificity of the Pap test. In North America, women with equivocal cellular changes or mild cytological abnormalities such as ASCUS and LSIL are required to return for a repeat Pap smear within 6 months. Women with two consecutive abnormal cytology results of ASCUS or LSIL are referred to colposcopy where a biopsy may be taken to determine if HSIL is present, requiring treatment [Ratnam et al., 2000; Solomon et al., 2001]. However, the majority of women are over-referred to colposcopy since only 1% of all

LSIL cases are believed to progress to HSIL or worse, based on results from a cohort study in Ontario with 17, 217 women [Holowaty et al., 1999; Miller et al., 2000]. The value of testing for HR-HPV types in the triage of women with ASCUS seems to be consistently superior to repeated cytology for the detection of HSIL as reviewed by previous studies [Unger & Duarte-Franco, 2001; Bosch et al., 2001]. It appears that significantly fewer women are unnecessarily referred to colposcopy when the decision is based on both a HR-HPV positive test and an ASCUS diagnosis [Manos et al., 1999].

However, HPV testing in the triage of LSIL cases is less consistent because of the high prevalence rates of HR-HPV DNA found in low-grade lesions in some studies. The ALTS (ASCUS & LSIL triage study) trial found that nearly all women with LSIL were HR-HPV positive and were referred to colposcopy immediately [The ALTS Group, 2000]. Nonetheless, while the positive predictive value of an HPV test limits its usefulness in the clinical management of LSILs, the negative predictive value may be an important benefit. The negative predictive value of a normal Pap smear result and a negative HPV test is nearly 100%. This observation could allow for extended screening intervals of 3-5 years instead of annually in a large proportion of women. Women older than 30 years of age would benefit the most from this strategy owing to the lower prevalence of infection in this age range [Ratnam et al., 2000; Meijer et al., 2000; Cuzick, 2000; Unger & Duarte-Franco, 2001].

2.1.9 Vaccination

Ultimately vaccination against HPV may have the greatest value, especially in developing countries. Prophylactic vaccines, currently in development are primarily based on DNA-free virus-like particles (VLP) synthesized by self-assembly of fusion proteins of the major HPV capsid antigen L1, and have been found to induce a strong humoral response with neutralizing antibodies [Harro et al., 2001]. Preliminary results from a large clinical trial with nearly 2400 young women (16-23 years of age) demonstrated that women in the HPV-16 VLP vaccine group were 100% protected from an incident HPV-16 infection and all nine cases of HPV-16-related cervical intraepithelial neoplasia occurred among the placebo recipients, over an average of 17

months follow-up [Koutsky et al., 2002]. However, one important issue needs to be resolved before the introduction of HPV vaccination in different populations. The humoral immunity against most HPV infections is type-specific [Roden et al., 1996; White et al., 1998]. Therefore, it will be of critical importance to determine how many HPV types, other than HPV-16, need to be targeted in the vaccine.

2.2 Pathogenesis of Cervical HPV Infection

2.2.1 Characterization of HPV

Papillomaviruses, members of the *Papillomaviridae* family, are widespread in nature inducing warts (or papillomas) primarily in higher vertebrates. [de Villiers, 1994]. The human papillomavirus (HPV) targets receptors in the basal epithelial cells and replicates in the nucleus of squamous epithelial cells. HPV genotypes are considered new when the nucleotide homology of specific regions of the viral genome (L1, E6 and E7 sequences) differs by more than 10%. In contrast, a subtype is specified as a type that exhibits between 90% and 98% homology, while HPV variants share less than 2% DNA sequence variation in their E6, E7 and L1 regions [de Villiers, 1994].

The viral genome of the papillomavirus can be divided into three regions (IARC 1995). The early region codes for proteins required for the regulation of viral transcription (E2), viral DNA replication (E1, E2), cell proliferation (E5, E6, E7) and possibly some late steps in the viral life cycle (E4). The late region contains two genes that code for the capsid proteins L1 and L2. The long control region (LCR) is a non-coding region that harbours the origin of replication and transcription control signals and is essential for regulatory functions of the genome [Koutsky et al., 1988].

The two major oncogenes identified in the HPV genome that encode proteins with growth stimulating functions are E6 and E7. The HR-HPV E6 proteins are also capable of inhibiting an important tumour-suppressor gene, p53 [Werness et al., 1990]. In stressed cells, p53 expression induced by DNA damage may result in cell cycle arrest, allowing for DNA repair before mutations are incorporated into newly synthesized DNA. In cases of extreme DNA damage, p53 may even trigger an apoptotic response [Lane, 1992]. *In*

vitro studies have demonstrated that E6 binding to p53 results in the rapid proteolytic degradation of the p53 proteins through an ubiquitin dependent pathway [Vousden, 1993]. This interference with p53 means reduced apoptosis and activation of endogenous proto-oncogenes including c-myc.

Similar to p53, the retinoblastoma gene (pRb) is another tumour suppressor gene known to be involved in the control of the G1 cell cycle, an important checkpoint for cell cycle regulation. Like E6 oncoproteins, HPV E7 proteins are involved in cellular transformation, capable of inhibiting pRb in the host [zur Hausen & Devilliers, 1994]. However, the affinity of the HR-HPV types to bind to the pRb are approximately ten-fold higher than the binding ability of LR-HPV types to the pRb. The observed association of E7 with cyclin A and cyclin dependent kinase cdk2 is a further indication that E7 expression may also stimulate cells to proceed through G2 phase to cell division. This interaction could disturb the normal control of entry into mitosis, regulated by cyclin dependent kinases, and may contribute to the development of HPV-associated cancers [Munger et al., 1989; Hamel et al., 1992; Vousden, 1993].

In LSILs, viral DNA is found in its episomal form in the nucleus of the squamous epithelial cells and there is low expression of the viral oncoproteins E6 and E7 [Clarke & Chetty, 2002]. In contrast, carcinoma *in situ* and invasive cervical cancer are characterized by high expression of viral oncoproteins E6 and E7. Integration of the viral genome into the host genome is a characteristic of transition from an HPV infection to cervical malignancy. This viral integration disrupts the E2 reading frame of the viral genome resulting in the loss of regulation of E6 and E7 expression [zur Hausen, 2000].

2.2.2 HPV type and cancer risk

Papillomaviruses are found exclusively in squamous cell epithelium. This specificity is further defined by their tropism for cutaneous squamous cell epithelium or for mucosal squamous cell epithelium. For the most part, HPV types are either mucosotropic or cutaneotropic, and are very specific to the anatomic site that they infect. Of the over 100 HPV types identified thus far, approximately 40 have been associated with the genital

tract (mucosotropic) [zur Hausen & Devilliers, 1994; IARC, 1995; zur Hausen, 2000]. Table 2.3 describes the most frequent genotypes found in cases with LSIL (CINI) and HSIL (CINII/III) reported in recent case-control and cohort studies. There is quite a lot of overlap between common types found in cervical lesions in different populations, although not one population has the exact same distribution of types detected in LSILs and HSILs. HPV 16, 31, 51, 53 and 58 appear to be very common in low-grade lesions while HPV 16 is almost always the most frequent type found in high-grade lesions followed by HPV 31, 52 and 58.

HPV types are usually divided into two groups based on their oncogenic potential. Some of the low risk types are common in low-grade squamous intraepithelial lesions (LSIL) but rarely are they found in high-grade SILs or invasive cancer. They include types 6, 11, 26, 32, 34, 40, 42, 44, 53, 54, 55, 57, 61, 62, 64, 66, 67, 69, 70, 71 (formerly CP8061), 72, 73, 81 (formerly CP8304), 82 (formerly MM4 orW13B), IS39 (subtype of 82), 83 (formerly MM7 or Pap291), 84 (formerly MM8 or Pap155), CP6108. The high risk types are strongly associated with cervical cancer and include types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68 [Bauer et al., 1993; Bosch et al., 2002].

However, the multicentre IARC study recently published results that suggested an additional two types (HPV 73 and HPV 82) that were originally considered low-risk should be considered as high-risk, based on pooled data from 11 case-control studies from nine countries [Munoz et al., 2003]. An additional three types (HPV 26, HPV 53 and HPV 66) were classified as probably high-risk.

This new classification may have implications for HPV screening, since hybrid capture II (HC-II), the only rapid HPV detection system to be currently approved by the FDA, just tests for the original panel of 13 suggested high-risk types [Bosch et al., 2002]. Moreover, comparisons of epidemiological studies that evaluate the natural history for HR- and LR-HPV infections separately will have to be made based on studies with the same epidemiologic classification for HPV types.

Table 2.3
The three most common HPV types found in women with low-grade squamous intraepithelial lesions (LSIL) and high grade squamous intraepithelial lesions (HSIL) as reported in seven studies

Study	Ranking of 3 most common HPV types found in women with LSIL in descending order of frequency (no. of cases of LSIL)	Ranking of 3 most common HPV types found in women with HSIL in descending order of frequency (no. of cases of HSIL)
[Herrero et al., 2000] -Costa Rica, N=9175	16 51 56/58 (n=181)	16 57 51/52 (n=125)
[Sasagawa et al., 2001]– Japan (cases)=366 (controls)=1,562	53 16 31/52 (n=145)	16 58 52 (n=137)
[Liaw et al., 1999]- USA. (cases)=380 (controls)=1037	16 51 53/56 (n=64)	16 18/59 6/35/39/52/56/58 (n=24)
[Ho et al., 1998]- USA N=258	31 58 16/52 (n=164)	16 52/58 31/56 (n=95)
[Schiff et al., 2000]- USA (cases)=112 (controls)=326	NA	31 16 58 (n=112)
[Moscicki et al., 2001]- USA N =496	16 18 31/33/35 (n=496)	NA
[Matsukura & Sugase, 2001], 2001 - Japan N=386	58 52 56 (n=98)	16 52 58 (n=188)

2.2.3 HPV transmission (viral entry)

While anogenital HPV infections are primarily sexually transmitted, infection of basal layer cells via a wound or an abrasion of a susceptible epithelial surface appears to play a major role in contracting HPV in the target epithelium [zur Hausen & Devilliers, 1994]. Junctions of different types of epithelial cells also appear to be highly vulnerable to HPV infection, possibly because of the proximity of proliferating cells to the surface of such sites. This junction, referred to as the transformation zone (T-zone), is where the stratified squamous epithelium of the ectocervix meets the mucus-secreting columnar epithelium of the endocervix. It is this area from which the majority of intraepithelial and invasive neoplasms of the cervix develop [Koutsky et al., 1988]. Therefore, when cervical lesions are identified (during colposcopic examination), the transformation zone is biopsied and histologically analyzed [de Roda Husman, 1995].

2.2.4 HPV DNA detection

Nucleic acid hybridization, based on the ability of DNA probes to hybridize with complementary sequences, offers a specific and sensitive approach for detection of DNA sequences. While both dot blot hybridization and Southern blot hybridization techniques were used extensively in early epidemiological studies of cervical HPV infection [Franco, 1991a; Brinton, 1992], the majority of studies in the last decade have used PCR or liquid hybridization (hybrid capture) methods to detect HPV [Kjaer et al., 1997; Schiffman et al., 2000].

Hybrid capture hybridizes denatured DNA to RNA probes in solution. The RNA-DNA complexes are transferred to a tube coated with an antibody directed against the RNA-DNA hybrids. An alkaline phosphatase conjugated antibody against RNA-DNA hybrids is added to the captured RNA-DNA hybrids. A chemiluminescent substrate then reacts with the alkaline phosphatase and the light that is produced is subsequently measured. A great advantage of this test is that it is quantitative and can be applied to crude cell suspensions. The quantity of produced light represents the amount of detected HPV DNA [de Roda Husman, 1995]. However, it is still a very costly method of analysis, but is the most obvious candidate to be employed for diagnostic purposes, should HPV testing as

an adjunct to cytology become routine in screening for cervical premalignant lesions [Schiffman & Schatzkin, 1994; Solomon et al., 2001].

PCR is an enzymatic *in vitro* DNA amplification reaction. DNA PCR consists of three steps that include DNA denaturation, annealing of primers (complementary oligonucleotides) that provide free 3'-OH ends for DNA polymerase mediated chain elongation in the last step. The process exploits a thermostable *Taq* polymerase allowing the steps to be repeated 30 to 40 times in the described order [Saiki et al., 1988]. Theoretically, a million fold increase in target DNA can be obtained resulting in a sensitivity of 1 HPV copy per sample. Very low amounts of input DNA (25-500 ng) are sufficient to generate detectable amounts of target DNA and this system can be easily applied to crude cell suspensions, eliminating the tedious task of DNA purification [Saiki et al., 1988; Bauer et al., 1991]. PCR can also be applied to detect target DNA in paraffin embedded tissue, although samples stored for prolonged periods of time (>7 years) have shown a marked reduction in PCR sensitivity [Greer et al., 1990]. Development of general mediated primers that amplify smaller fragments of HPV DNA have greatly improved the PCR sensitivity for HPV detection in paraffin-embedded tissue [Walboomers et al., 1994; Giannoudis & Herrington, 2001]. The combination of PCR and DNA hybridization has gradually become the gold standard in HPV detection techniques. However, PCR's greatest strength is also its greatest liability. Its superior sensitivity gives rise to increased risk of laboratory contaminations and excessive laboratory anti-contamination measures must be followed to ensure reliable results [Bauer et al., 1993].

2.3 Epidemiology of Cervical HPV Infection

2.3.1 Occurrence of cervical HPV infection

The epidemiological study of cervical HPV infection has largely been determined by available methods for diagnosing and defining infection [Schiffman & Schatzkin, 1994]. Some inconsistencies emerged in early molecular epidemiology studies of HPV and cervical cancer that used first generation DNA hybridization methods to detect the virus. Contrary to expectations, these epidemiological studies were not able to find an association between cervical HPV infection and predictors of sexual activity [Villa &

Franco, 1989; Reeves et al., 1989; Kjaer et al., 1990]. These paradoxical findings, considering that cervical cancer risk is strongly associated with sexual behaviour, were subsequently explained by measurement errors of HPV status [Franco, 1991a; Franco, 1992]. Modern PCR protocols based on consensus primers have become the preferred methods for epidemiological studies in recent years, because of their high sensitivity and specificity [Manos et al., 1989; van den Brule et al., 1991], while the HC-II system is the preferred diagnostic tool for HPV screening trials[The ALTS Group, 2000].

2.3.1.1 Prevalence of HPV infection

The majority of epidemiological studies using PCR have found the prevalence of HPV infections to vary from 10% to 45%, peaking in women 15-25 years of age with a second peak recently observed in women over the age of 50 [Sellors et al., 2000; Herrero et al., 2000; Lazcano-Ponce et al., 2001]. Colposcopic examinations were not done in the majority of these studies, so there is no evidence on how many patients had actual cervical lesions. Therefore, most of the reported point prevalence estimates comprise subclinical and latent disease. Table 2.4 is a summary of current worldwide HPV DNA detection rates by PCR among women with normal cytology, attending cervical screening programs.

Prevalence of HPV varies greatly in different populations, dependent on both demographic and behavioural determinants of infection (table 2.4). In particular, young, sexually active women (less then 25 years of age) appear to experience the highest HPV prevalence, ranging between 18%-46% [Ley et al., 1991; Wheeler et al., 1993; Kjaer et al.,1993; Herrero et al., 2000; Sellors et al., 2000; Lazcano-Ponce et al., 2001]. Older monogamous women, over 35 years of age are much less likely to be HPV positive [Munoz et al., 1993; Melkert et al., 1993] [Franco et al., 1995] with the background HPV prevalence level decreasing to 2-8% in most populations [Bosch et al., 2002].

Differing estimates may be attributed in part to the age distribution and risk factors for genital infection in the populations examined and by the variation in sample collection (cervico- vaginal lavage [collection of cervical and vaginal exfoliated cells], spatula [usually collects ectocervical cells], cervical brush [usually collects endocervical cells],

or cervex brush [collects both ecto- & endocervical cells]). In addition, even when restricting the comparison to studies using PCR methods, the possibility of inter-laboratory variation cannot be ignored [Schiffman, 1992]. A number of changes in HPV DNA detection with PCR have occurred since its introduction in the late 1980's [Manos et al., 1989]. Primers have been re-designed to minimize base-pair mismatches so that amplification of certain HPV types has improved [Gravitt et al., 2000]. In addition, some investigators have adapted their MY09/MY11 PCR protocols with the introduction of a new polymerase, Amplitaq Gold (TaqGold; Perkin-Elmer Cetus, Norwalk, CT), since it is believed to have better enzymatic characteristics for PCR reactions. A recent study demonstrated that the use of Amplitaq Gold with MY09/MY11 primers allowed for the preferential detection of HPV DNA positive specimens with lower viral load than the MY09/MY11 protocol using an earlier generation of *Taq* polymerase [Castle et al., 2002]. Thus, in describing HPV prevalence and the distribution of HPV infections worldwide, the population profile and the method of HPV DNA detection and sample collection should also be considered.

Table 2.4

HPV DNA prevalence among women with normal cervical cytology

Study	Country	HPV + (%)	Population profile	Number of HPV types tested & detection system used	Method of sample collection
Van den Brule *et al.*, 1990	Holland	25.0	Hospital	GP[1] &TS[2] PCR: types 6, 11, 16, 18, 31, 33,	Cervical scrape
Hallam *et al.*, 1991	England	56.0	Family clinic	MY09/11 PCR: types 6, 11, 16, 18	Ayre spatula
Ley *et al.*, 1991	USA	46.0	University students (18-30 yrs)	MY09/11 PCR: types 6, 11, 16, 18, 31, 33	Cervical swab
Rohan *et al.*, 1991	Canada	18.1	University students (18-30 yrs)	MY09/11 PCR: types 6, 11, 16, 18, 31, 33	Ayre spatula
Vandenvelde & Vanbeers, 1992	Belgium	16.6	Screening clinic	TS PCR: types 6, 11, 16, 18	Cervical scrape
Bauer, 1993	USA	17.7	Family planning clinic	MY09/11 PCR: ~25 types, HR[3] (12)	Cervico-vaginal lavage
Bosch *et al.*, 1993 (Controls)	Spain Colombia	4.7 10.5	Family clinics Pap screenees	MY09/11 PCR: types 6, 11, 16, 18, 31, 33	Ayre spatula & cytobrush
Hildesheim *et al.*, 1993	USA (DC)	33.7	Medical assistance clinics	MY09/11 PCR: ~25 types, HR (12)	Cervico-vaginal lavage
Kjaer *et al.*, 1993	Greenland Denmark	43.4 38.9	General population	GP PCR: types 6, 11, 16, 18, 31, 33	Cervical swab
Wheeler *et al.*, 1993	USA (NM)	44.3	University students (18-30 yrs)	MY09/11 PCR: ~20 types, HR (9)	Cervical swab
Franco *et al.*, 1995	Brazil	18.3	Family planning clinic (18-60)	MY09/11: ~40 types HR (13)	Ayre spatula & cervical brush

[1]GP=general primers; [2]TS=type-specific primers; [3]HR=number of HR-HPV types tested

Table 2.4 (continued)

Study	Country	HPV + (%)	Population profile	Number of HPV types tested & detection system used	Method of sample collection
Burk *et al* 1996	USA (NY)	27.8	University students	MY09/11: ~40 types HR (16)	Cervico-vaginal lavage
Giuliano *et al* 1999	USA (AZ)	13.2	County health clinic	HC tube system HR (9)	Cervical scrape
Richardson *et al* 2000	Canada (QC)	21.8	University students (18-30 yrs)	MY09/11: ~40 types HR (13)	Ayre spatula & cervical brush
Herrero *et al* 2000	Costa Rica	16.0	General population (18-94 years)	MY09/11: ~40 types HR (13)	Cervex brush
Lazcano-Ponce *et al* 2001	Mexico	12.8	General population (18-65+)	MY09/11: 27 types HR (17)	Ayre spatula & cervical brush
Sellors *et al* 2000	Canada (ON)	12.7	Family clinics (15-49	HC-II: HR (13)	Cervical brush
Peyton *et al* 2001	USA (NM)	39.2	Women's health clinics (18-40 yrs)	MY09/11 & Taq-Gold: ~40 types HR (18)	Cervical swab
Lüdicke *et al* 2001	Switzer-land	14.2	Adolescent clinic (14-20 yrs)	HCII: HR (13)	Cervex brush
Chan *et al* 2002	Hong-Kong	7.3	Screening clinic	MY09/11: ~20 types HR (9)	Cervex brush

[1]GP=general primers; [2]TS=type-specific primers; [3]HR=number of HR-HPV types tested

The peak in cervical HPV prevalence in 15-25 year olds in North America corresponds to the usual age of initiation of sexual intercourse. The prevalence of HPV drops dramatically in women over 30 years of age and might be due to immunologic clearance or suppression of existing infections. Alternatively, it may be explained by less exposure to new HPV types because of fewer new sexual partners [Schiffman & Schatzkin, 1994].

While the prevalence of cervical HPV infections has been investigated in numerous studies [IARC, 1995], the geographical variation in type distribution has not been extensively documented, except for HPV 16, which appears to be the most frequently occurring type in most countries [Ley et al., 1991; Koutsky et al., 1992; Wheeler et al., 1993; Hildesheim et al., 1994; Ho et al., 1998; Ho et al., 1998; Liaw et al., 1999; Franco et al., 1999; Liaw et al., 1999; Franco et al., 1999; Woodman et al., 2001; Woodman et al., 2001]. HPV 53 and 51 appear to be very prevalent in different populations but the high prevalence of HPV 84 (previously MM8 or Pap155) seen in younger women [Ho et al., 1998; Richardson et al., 2000; Giuliano et al., 2002a] has not been seen in cohorts of middle-aged women, in which HPV 84 is very rare [Liaw et al., 1999; Franco et al., 1999; Liaw et al., 1999; Franco et al., 1999; Lazcano-Ponce et al., 2001].

When HPV types are grouped according to their oncogenic potential, the distribution of prevalent HR-and LR-HPV infections appears to vary between different studies. Explanation for the discrepancy between studies could include differences in classification of HR- and LR-HPV types. In addition, some groups tested for fewer low-risk types than other groups. However, the question of whether high-risk types occur more frequently or are just more persistent, and therefore appear more prevalent, can only be resolved by studying the incidence of HPV genotypes in prospective cohort studies.

2.3.1.2 Incidence of HPV infections

The incidence of HPV infections has been described for a variety of cohorts, although the majority of studies have been conducted among young women (<35 years) attending university or community clinics. The incidence has either been presented as the cumulative risk of acquiring a new HPV infection over a specified amount of time (calculated with actuarial techniques) or expressed as the incidence rate using person time in the denominator. Table 2.5 is a summary of the few prospective cohort studies that were able to calculate HPV incidence. The 3-year cumulative incidence for acquiring a new HPV infection appears to range between 43% and 55% [Ho et al., 1998; Franco et al., 1999; Woodman et al., 2001; Moscicki et al., 2001]. Three of these studies were very similar with regards to population and age structure. Ho et al. [Ho et al., 1998] followed

University students from New Jersey (mean age 20 years), Moscicki et al. [Moscicki et al., 2001] recruited young women from a University Health Clinic (mean age 21) and a Planned Parenthood clinic (mean age 18), both located in San Francisco and Woodman et al. [Woodman et al., 2001] assembled a cohort of women between 15-19 from a community clinic in England. Only Franco et al. [Franco et al., 1999] studied an older cohort of women from Brazil who ranged in age from 18 to 69 (mean age 33 years). Nonetheless the estimated cumulative incidence of 44% [Rousseau, 2003] was very similar to the cumulative probabilities observed in the three younger cohorts in the US and Britain. Two other studies based in the United States have observed slightly higher cumulative probabilities for HPV infection. The 2-year cumulative incidence for any new HPV infection among a cohort of women attending college in Washington was 39% [Winer et al., 2003] while a 1-year cumulative incidence of 40% was observed among a cohort of 18-35 year old women attending planned parenthood clinics in Arizona [Giuliano et al., 2002a].

Of the studies that estimated the cumulative incidence for HR- and LR-HPV infections separately (table 2.5), two observed that LR-HPV types were acquired less frequently than HR-HPV types [Moscicki et al., 2001; Giuliano et al., 2002a] while the opposite was observed in another study [Franco et al., 1999]. However, Giuliano et al. [Giuliano et al., 2002a] preselected their cohort based on HR-HPV positive test results at baseline and Moscicki et al. [Moscicki et al., 2001] only tested for half a dozen low risk types. Franco et al. [Franco et al., 1999] tested for all 40 individual HPV types, of which the vast majority are low-risk types, which may explain the discrepancy in these results.

The incidence rate for any HPV infection is summarized in table 2.5, and appears to range from 16 new cases per 100 woman-years [Franco et al., 1999] to 47 cases per 100 woman-years among a cohort of women attending an STD clinic in Amsterdam [Van Doornum et al., 1994]. The rate of HPV acquisition among a cohort of university students in Seattle [Thomas et al., 2000] and a second cohort from Arizona was 28 and 35 per 100 woman-years [Giuliano et al., 2002a], respectively. The difference in incidence rates between the various cohorts suggests that age is the strongest predictor of HPV

acquisition, most likely because it can be considered to be a proxy for HPV exposure after the first initiation of sexual intercourse [Schiffman & Brinton, 1995]. Another factor that may also have contributed to the different incidence rates is the varying levels of high-risk sexual behaviour among women in the different cohorts. The rate of acquiring any new LR-HPV infection was slightly higher than that for acquiring any new HR-HPV infection among women who were HIV negative but had a history of sexual or drug abuse [Ahdieh et al., 2001]. Franco and collaborators observed similar results in a lower-risk cohort in Brazil [Franco et al., 1999].

Table 2.5
Cervical HPV incidence rates and cumulative incidence among different cohorts

Author	Incidence rate	Cumulative incidence (95% CI)		
	Per 100 woman-years (95% CI)	Overall HPV	HR-HPV	LR-HPV
Van Doornum, 1994	47.1 (33.8,60.3)	*N.A.	N.A.	N.A.
Ho et al, 1998	N.A.	*3 year:* 43% (36, 49)	N.A.	N.A.
Franco et al, 1999	Any: 16.1 (13.7,18.8) **HR: 8.2 (6.5,9.8) ***LR: 10.8 (9.0,13.1)	*18 mo:* 24% (17, 30) *36 mo:* 44% (36, 51)	*18 mo:* 11% (7, 16)	*18 mo:* 22% (15, 30)
Thomas et al, 2000	27.8 (N.A.)	N.A.	N.A.	N.A.
Moscicki et al, 2001	N.A.	*1 year:* 17% (14, 19) *3 year:* 55% (31, 79)	*1 year:* 28% (N.A.)	*1 year:* 4.6% (N.A.)
Woodman et al, 2001	N.A.	*3 year:* 44% (40, 48)	N.A.	N.A.
Ahdieh et al, 2001	HR: 4.4 (N.A.) LR: 7.4 (N.A.)	N.A.	N.A.	N.A.
Giuliano et al, 2002	Any: 35.3 (24.7,48.8)	*1 year:* 40% (N.A.)	*1 year:* 32% (N.A.)	*1 year:* 18% (N.A.)
Winer, 2003	N.A.	*2 year:* 39% (33, 45)	N.A.	N.A.

**N.A: Not available; **HR: High risk HPV infections; ***LR; Low risk HPV infections*

2.3.1.3 Multiple-type HPV infections

Co-infection with other HPV types at the same visit (concurrent co-infection) or at sequential visits (sequential co-infection) is extremely common, particularly in young women. Of those women who are HPV positive at a given visit, the proportion of cumulative co-infections (concurrent or sequential) can vary between 11% to 45% [Bauer

et al., 1993; Rousseau et al., 2000; Ahdieh et al., 2001; Moscicki et al., 2001; Woodman et al., 2001]. Age seems to be a strong determinant of multiple-type HPV infections, reflected in the high proportion of co-infections (30-45%) in young cohorts where the mean age was less than 27 years-old [Wheeler et al., 1993; Hildesheim et al., 1994] and the lower proportion of co-infections (<17%) in the older cohorts where the mean age was greater than 32 years [Bauer et al., 1993; Ho et al., 1998; Herrero et al., 2000]. Thomas et al [Thomas et al., 2000] concluded that women were more likely to experience co-infection with multiple types than would be expected by chance alone. A co-infection was associated with HPV persistence at the subsequent visit in one cohort of university students [Ho et al., 1998] and Ahdieh et al. [Ahdieh et al., 2001] observed an elevated risk of HPV persistence among women with co-infections who were also positive for HIV. However, other studies, among lower-risk populations, have not always observed an increased risk of persistence associated with co-infections [Liaw et al., 1999; Rousseau et al., 2001].

2.3.2 Duration of cervical HPV infections

The definition of HPV persistence has varied significantly between studies designed to estimate the average duration of HPV infections and to elucidate the relevant risk factors for persistent HPV infections, among women with normal cervical cytology. Some studies have based HPV persistence on sequential pairs of visits that were positive for HPV over 3 or more years of follow-up [Ho et al., 1995; Ho et al., 1998; Ahdieh et al., 2001] while other studies have defined persistence based on two visits [Hildesheim et al., 1994; Brisson et al., 1996; Sedjo et al., 2002b] or on time-to-HPV clearance [Moscicki et al., 1998]. Studies using PCR or Southern blot as their HPV detection system have based HPV persistence on the detection of the same type(s) at two consecutive visits [Hildesheim et al., 1994; Ho et al., 1998; Ahdieh et al., 2001] or the same oncogenic HPV group at 2 or more visits [Moscicki et al., 1998; Franco et al., 1999; Schlecht et al., 2001; Sedjo et al., 2002a; Giuliano et al., 2002a] or the detection of any HPV type at two consecutive visits [Brisson et al., 1996; Woodman et al., 2001]. Because the majority of studies have not assessed type-specific persistence, the number of persistent cases may

have been overestimated by including subjects whose original infection was replaced by a newly acquired infection of a different HPV type.

Table 2.6 describes the results from natural history studies of cervical HPV infection that included repeated sampling of women for viral persistence. The median duration of new or prevalent HPV infections ranges between 8 months and 14 months in young [Ho et al., 1998; Woodman et al., 2001] [Giuliano et al., 2002a] and middle aged women [Franco et al., 1999]. However, distinguishing between an infection that has truly cleared and a false negative test result due to poor sampling, low levels of virus or insensitive measurement tests is very difficult [Woodman et al., 2001]. Consequently the clearance rate may be somewhat over-estimated in some of these studies, while the frequency of persistent infections may be underestimated.

Moscicki and collaborators [Moscicki et al., 1998] considered various definitions of clearance by modeling different number of consecutive HPV negative tests since the last HPV positive test. Expectedly, as the definition of clearance became more conservative, the median duration of the infection increased. The authors concluded that it took 20 to 30 months for the majority of women in their study to clear a prevalent HPV infection, conditional on three consecutive HPV negative tests. However, this research group did not use a PCR amplification system for HPV detection when they initiated their study in 1990, because the MY09/11 PCR system was still in development [Manos et al., 1989]. While many of the samples were re-tested in the mid to late 1990's with PCR, many were not yet re-evaluated at the time that the results on clearance rates were published [Moscicki et al., 1998]. The lack of a highly sensitive HPV detection system may explain the long duration observed by this group.

Table 2.6

Median duration and/or clearance rates for HPV among different cohorts

Author (mean age of cohort members)	Median duration of HPV infection in months (95% C.I.)	Proportion of women who cleared HPV	Median time between visits
Hildesheim et al, 1994 (26 years)	*N.A.	*~1 year* 37%	14.6 mo. (2 visits only)
Evander et al, 1995	N.A.	*2 years* 95%	N.A. (2 visits only)
Brisson et al, 1996 (~29 years)	N.A.	*~6 months* 50%	3.5 mo. (2 visits only)
Ho et al, 1998 (20 years)	8 (7,10)	N.A.	6 mo. (multiple repeat visits)
Moscicki et al, 1998 (~20 years) *(after 3 HPV- visits)*	Any HPV: ~15 (N.A.) **HR: ~12 (N.A.) ***LR: ~10 (N.A.)	*2 years* 70%	4 mo. (multiple repeat visits)
Franco et al, 1999 (33 years)	HR: 8.1 (7.8, 8.3) LR: 4.8 (3.9, 5.6)	N.A.	4 mo. (multiple repeat visits)
Elfgren et al, 2000	N.A.	*5 years* 92%	5 years (2 visits only)
Woodman et al, 2001 (18 years)	13.7 (8.0-25.4)	N.A.	N.A. (multiple repeat visits)
Ahdieh et al, 2001 (~30 years)	Type-specific: ~6 (N.A.)	*3 years* HR: 91.5% LR: 87.5%	6 mo. (multiple repeat visits)
Giuliano et al, 2002 (~28 years)	Any HPV: ~9 (N.A.) HR: 9.8 (N.A.) LR: 4.3 (N.A.)		~5 mo. (3 visits)
Sellors et al, 2003 (33 years)		*~1 year 51.9%*	14.0 mo. *(only 44% f-up rate)*

**N.A: Not available; **HR: High-risk HPV infections; ***LR: Low-risk HPV infections*

One study noted that the median duration of prevalent low-risk infections was about 5 months compared with 8 months for high-risk types (table 2.6). However, when the mean infection time was estimated instead, low-risk infections endured for an average of 8 months, while high-risk infections persisted for nearly 14 months [Franco et al., 1999]. The mean duration time is more influenced by persistent infections than the median duration time, which is a measure of the time that 50% of the women cleared their first episode of an HPV infection. Approximately 80% of women infected with HPV are estimated to clear their infection before a lesion develops [Ho et al., 1995; Nobbenhuis et al., 1999]. Therefore, given that only 20% of the women are estimated to harbour persistent HPV infections, the median duration may be under-estimating the true average duration of transient infections.

Nonetheless, while it is generally agreed that the majority of HPV infections are transient, there is tremendous variation in the proportion of type-specific HPV persistent infections observed in the literature. Among studies in which 2 or more visits were considered in the determination of HPV status, the proportion of HPV persistent infections in a one year period has been observed to range between 35% [Franco et al., 1999] to over 50% [Ho et al., 1995; Giuliano et al., 2002a; Sellors et al., 2003] or 60% [Hildesheim et al., 1994]. One important factor that could influence the assessment of persistence is the time interval between visits [Brisson et al., 1996]. With short testing intervals, such as every 3 months, the proportion of persistent infections at two consecutive visits would appear greater than would be observed if the testing intervals were 6 months apart. The alternate scenario can occur when the testing interval is longer and an apparently persistent HPV infection may actually be a new HPV infection [Hildesheim et al., 1994].

Defining persistence based on the same HPV type observed at repeated visits improves the classification of persistence, but only variant analysis of specific HPV types can really determine a new from persistent infection. Variant analysis for incident HPV 16 infections was conducted among a cohort of university students in Washington who returned every 4 months for a follow-up visit. The median duration of an incident

prototype (European) variant was 17 months and 14 months for those women with the less frequent Non-European (non-prototype) variant [Xi et al., 2002].

2.3.3 Risk factors for prevalent or incident HPV infections

2.3.3.1 Age

The most important determinant for HPV infection is age, with most studies indicating a sharp decrease in prevalence after age 30. While age has been shown to be strongly and positively associated with increasing lifetime number of sexual partners [Ley et al., 1991], the decrease in HPV infection risk with increasing age has also been shown to be independent of sexual activity [Wheeler et al., 1993] and, at least in certain populations, restricted to low oncogenic risk types [Franco et al., 1995]. The decrease of HPV prevalence with age could be suggestive that many HPV infections are transient, reflecting a mounted immune response. The recent observations of increased HPV prevalence among women over 50 years of age [Herrero et al., 2000; Lazcano-Ponce et al., 2001], paradoxically supports this hypothesis. Aging is associated with lowered immunity, potentially facilitating the re-activation of latent (but undetectable) HPV infections.

2.3.3.2 Sexual activity

The other key determinants of any prevalent HPV infection have consistently been age at first intercourse and number of lifetime partners [Ley et al., 1991] [Hildesheim et al., 1993; Wheeler et al., 1993] [Melkert et al., 1993; Bosch et al., 1994] [Kruger-Kjaer et al., 1998]. Furthermore, strong evidence from 5 studies with virgins showed that no HPV DNA was detected [Fairley et al., 1992; Anderson-Ellstrøm et al., 1996] or less than 2% of virgins had HPV detected [Rylander et al., 1994; Kruger-Kjaer et al., 2001; Winer et al., 2003]. Only those virgins that initiated sexual activity became positive for HPV DNA and/or seropositive for HPV 16 [Kruger-Kjaer et al., 2001; Winer et al., 2003].

However, some studies have found that the association of markers of sexual activity and prevalent HPV infection varied according to the oncogenic potential of the HPV type [Rousseau et al., 2000; Richardson et al., 2000] [Franco et al., 1995; Kjaer et al., 1997]

[Chan et al., 2002]. Infection with low oncogenic risk types was only weakly associated with lifetime sexual behaviour, whereas variables of lifetime sexual activity were much stronger predictors of high oncogenic risk HPV types, regardless of age [Franco et al., 1995; Richardson et al., 2000; Chan et al., 2002].

To date, there are far fewer cohort studies that have studied the determinants of incident HPV infection and amongst these a variety of different predictors of HPV transmission have been identified. The majority of studies found that age at first intercourse was not an independent predictor of incident HPV infection and recent sexual behavior was a stronger predictor of acquisition than number of lifetime partners[Ho et al., 1998; Elfgren et al., 2000; Kruger-Kjaer et al., 2001; Moscicki et al., 2001; Winer et al., 2003].

A woman's risk can also be increased by the sexual behaviour of her regular male partner. The male partner's sexual behaviour (particularly in populations where female monogamy is culturally dominant) has been considered a central determinant of the incidence of cervical cancer for over 30 years [Bosch et al., 2002]. Recently, Ho et al [Ho et al., 1998] and Winer et al. [Winer et al., 2003] both observed an increased risk for acquisition of a new HPV infection among women whose partners reported a higher number of lifetime sexual partners.

2.3.3.3 Non-sexual transmission

However, nonsexual transmission of HPV cannot be entirely ruled out. Studies have reported both perinatal transmission of HPV from mother to child, [Sedlacek et al., 1989; Pakarian et al., 1994] and presence of HPV in the foreskin from newborns [Roman & Fife, 1986] as well as in the oral mucosa of healthy preschool children [Jenison et al., 1990]. In addition, HPV DNA may have the ability to adhere to certain materials, facilitating transmission to the cervix, based on findings of HPV DNA on underclothes and gynaecological equipment [Ferenczy et al., 1989; Ferenczy et al., 1990].

2.3.3.4 Parity and oral contraceptives

Acquisition of HPV infection may also be influenced by other variables, such as parity and oral contraceptive use. Multiple full-term births (multiparity) may facilitate the

acquisition of persistent HPV infections by providing cumulative opportunities for immunosuppression and breaches in the cervical epithelium [Franco, 1993]. It has been shown that genital condylomas, tend to increase in size during the later stages of pregnancy [Koutsky et al., 1988] and that the detection rate of asymptomatic HPV infection of the cervix tends to be higher in pregnant women [Schneider et al., 1987; Rando et al., 1989; Sethi et al., 1998]. There is also evidence to suggest that endogenous sex steroids and exogenous hormones, such as oral contraceptives, may affect a woman's susceptibility of contracting a genital HPV infection [Brabin, 2002]. While little is known about mucosal immunity to HPV [Tjiong et al., 2001], the presence of IgA and IgG antibodies against HPV have been detected in cervical secretions and serum. Sex steroids appear to affect mucosal immunity and may exert stage-specific effects on HPV infection, whereby the risk of HPV acquisition is reduced with increased levels of sex steroids; but once an HPV infection is established, high levels of estrogen may increase likelihood of persistence [Brabin, 2002].

Current oral contraceptive use was found to be significantly protective against the acquisition of an incident HPV infection in a recent cohort study of young women [Moscicki et al., 2001] and appeared to be protective for both oncogenic and nononcogenic (prevalent) HPV infections in a cross-sectional study of Montreal university students [Richardson et al., 2000]. Furthermore, women who were using hormone contraceptives for 4 or more years were half as likely to have a prevalent HPV infection than women who were not using oral contraceptives, among a group of women of reproductive age in Arizona [Giuliano et al., 1999].

However, a woman's hormonal status is also affected by a range of other endogenous and exogenous hormonal factors such as age, pregnancy, illness, drug therapies and possibly smoking [Brabin, 2002]. Therefore, older women using oral contraceptives may not have a lower risk of acquiring an HPV infection compared to younger (adolescent) women. Oral contraceptive use [Hildesheim et al., 1993; Bauer et al., 1993], the length of OC use [Brisson et al., 1996; Rousseau et al., 2000] and older age (>19 years) at first use of OC's [Kjaer et al., 1990] were associated with the presence of HPV in 5 studies with middle-

aged women. Still other cohort studies with either young or middle-aged women have not observed any association between oral contraceptive use and HPV infection [Ho et al., 1998; Elfgren et al., 2000; Kruger-Kjaer et al., 2001]. Whether oral contraceptives preferentially hinder HPV acquisition in young women by increasing levels of estrogen that boost mucosal immunity, or facilitate the acquisition of HPV in older women by altering the cervical epithelium by some unknown mechanism has yet to be confirmed.

2.3.3.5 Tobacco

The association between smoking and risk of a prevalent or new HPV infection has also been mixed, in the literature. Results from cross-sectional and cohort studies show smoking to be either a risk factor for overall HPV infections [Bauer et al., 1993; Winer et al., 2003], LR-HPV infections [Rousseau et al., 2000], HR-HPV infections [Chan et al., 2002]or not associated with HPV at all, after adjusting for markers of sexual activity and age [Ho et al., 1998; Deacon et al., 2000; Ludicke et al., 2001; Moscicki et al., 2001; Giuliano et al., 2002b]. However, the majority of studies that have observed an association with tobacco use and HPV infection were based on cross-sectional data [Bauer et al., 1993; Chan et al., 2002] or data that included prevalent cases at enrollment [Rousseau et al., 2000]. Therefore tobacco use may be acting as a co-factor that can influence HPV persistence, rather than acting as a facilitator for HPV acquisition.

2.3.3.6 Barrier methods of contraception

The protective effect of condoms has still not been confirmed. Observations have been inconsistent in both cross-sectional and cohort studies. Protective effects of condom use after adjusting for sexual activity were observed in some studies for overall HPV [Kotloff et al., 1998; Deacon et al., 2000], HR-HPV [Kjaer et al., 1997; Richardson et al., 2000] and LR-HPV infections [Chan et al., 2002] but not in other studies of overall HPV [McNicol & Young, ; Thiry et al., 1993; Ho et al., 1998; Ludicke et al., 2001; Moscicki et al., 2001; Winer et al., 2003], after controlling for sexual activity. Assessing the effect of condom use and HPV infection is very difficult, not least of all because of the high potential for bias reporting. Reporting of condom use may be particularly biased if women are using condoms for contraceptive purposes but not necessarily as a barrier against sexually transmitted diseases. In which case, the condom might not be used at all

times during sexual intercourse, allowing the chance of vaginal contact with HPV or other infectious agents. Adding to the complexity of the situation, studies that stratified HPV based on their oncogenic potential observed condom use to be protective for HR-HPV infections but a risk factor for LR-HPV infections, after adjusting for markers of sexual activity [Kjaer et al., 1997; Richardson et al., 2000]. Whether this means that genital HPV types have altered biological tropisms and can be found, differentially, on specific anatomical genital sites that are not all protected by a condom still needs to be clarified [Richardson et al., 2000; Ludicke et al., 2001].

2.3.3.7 Other sexually transmitted diseases

The role of other sexually transmitted infections and the acquisition of HPV are still not clear. Human immunodeficiency virus (HIV) infection has been associated with a very high prevalence of HPV DNA, [Vermund et al., 1991; Ho et al., 1994; Moscicki et al., 2001] and the incidence of HPV infection has been as high as 95% among HIV positive women [Clarke & Chetty, 2002]. There is also evidence that HSV has oncogenic properties [Aurelian et al., 1989; Franco, 1991b], and a synergistic interaction between HPV and HSV may exist in cervical carcinogenesis. Women with a history of HSV or venereal warts were more likely to acquire an HPV infection among a young cohort of women in California (Moscicki 2001), but these results have not been observed in other prospective studies. Past or current *Chlamydia trachomatis* infections have been associated with an increased likelihood of HR-HPV DNA detection in cross-sectional studies [Kjaer et al., 1997; Giuliano et al., 2002b] and a recent study noted that cervical inflammation was associated with an increased risk for high-grade neoplasia among women infected with HPV [Castle et al. 2001]. However, it is unclear whether these infections would act to irritate the cervical epithelium, thereby facilitating HPV transmission, or whether they would act as co-factors that promote HPV pathogenesis and persistence.

2.3.3.8 Hygiene practices

There are very few published reports describing the relationship between specific aspects of hygiene and HPV infection or cervical cancer [Brinton et al., 1987]. Of those studies that have enquired about hygiene practices, varied risks for HPV infection were observed

among women in different populations with different modes of personal hygiene. The practice of douching has been shown to be protective against HPV infections in at least one study [Rousseau et al., 2000] but not others [Ho et al., 1998; Richardson et al., 2000]. Washing after sex was shown to be significantly protective for prevalent HPV infections in one study [Richardson et al., 2000] and the use of tampons instead of sanitary napkins during menstruation has been associated with an elevated risk for cumulative HPV infections [Rousseau et al., 2000].

2.3.3.9 Caveat

There are two issues to consider when evaluating the inconsistent evidence on risk factors for HPV infection. The first consideration is that the vast majority of studies have only investigated determinants of prevalent HPV infection. Prevalent HPV infections are a mix of transient and persistent infections, although the proportion of persistent infections will vary according to the population and available screening for those individuals. Nonetheless, it is very possible that certain risk factors identified in cross-sectional studies may be predictors of persistent infections rather than incident infections. Secondly, risk factors for HPV transmission other than sexual activity may differ geographically, depending on the distribution of other risk factors in different populations. For example, perhaps the full influence of condom use on HPV risk can only be observed in regions where consistent and regular condom use is prevalent, and the weak effects observed in so many studies is a reflection of the low prevalence of regular use.

2.3.4 Risk factors for persistent HPV infection

There are numerous strategies available for defining persistence, as discussed in an earlier section on the duration of HPV infections. Therefore it is difficult to compare results from different studies that have evaluated the relationship between potential risk factors and HPV persistence. Nonetheless, some predictors have consistently been identified in the literature, regardless of study design and will be discussed in more detail in the next section.

2.3.4.1 Virus and host characteristics

The most important determinants of HPV persistence, thus far, appear to be factors related to the virus (viral type, viral load, co-infection) and host (genetic susceptibility, immunosuppression) [Ahdieh et al., 2001; Bosch et al., 2002; Clarke & Chetty, 2002]. HPV persistence is strongly associated with HR-HPV types [de Sanjose et al., 1994; Moscicki et al., 1998; Ho et al., 1998], particularly HPV 16 [Hildesheim et al., 1994; Elfgren et al., 2000]. However, certain variants of HPV 16 [Londesborough et al., 1996] or HPV 18 [Villa et al., 2000] may prove to be responsible for the increased probability of persistence associated with HPV 16 or -18 infections.

Each HPV genotype can be classified into variants defined by a limited number of genomic variations in coding and non-coding regions of the HPV. HPV 16 evolved along 5 major geographic branches and are often classified into two main groups of HPV 16 variants; the European group which contains the HPV 16 prototype and other variants that only vary from the prototype by 2 or less base pairs within a particular region of the HPV-E6 gene and the Non-European (NE) group (including the African, Asian and Asian American subgroups) [Ho, 1991; Xi et al., 1997].

Several researchers have found the HPV 16 Non-European variants to be associated with cervical cancer and lesions [Xi et al., 1997; Villa et al., 2000], and HPV persistence more often than the European variants. Common to both groups of HPV 16 variants is an additional variant associated with a base change in the E6 coding region from nucleotide 350T to nucleotide 350G, which results in an amino acid change from leucine to valine. The nucleotide 350G variant has been shown to be associated with HPV 16 persistence and CIN persistence [Yamada et al., 1995; Londesborough et al., 1996].

Viral load has also been recognized as a possible predictor of more severe and persistent HPV infections. However, the methods for quantifying viral load have varied from "guestimating" the quantity from the intensity of an unamplified signal on an electrophoresis gel to employing sophisticated quantitative PCR techniques. The majority of studies evaluating viral load have focused on it as a predictor of SIL development or progression but a few studies have also studied the influence of viral load on HPV

clearance or persistence. Women with the highest tertile for viral load, quantified with HC-II, were significantly less likely to clear their HPV infection than women in the lowest tertile, among a cohort of women with normal cytology in Amsterdam [Rozendaal et al., 2000]. Similarly, increased signal intensity on an ethidium-bromide stained-gel before [Brisson et al., 1996] or after PCR amplification [Ahdieh et al., 2001], suggestive of a higher viral load, was associated with HPV positivity at 2 visits (10 week interval, on average) [Brisson et al., 1996] and multiple consecutive visits (6month intervals, on average) [Ahdieh et al., 2001].

A co-infection with two or more HPV types at the same visit, or cumulatively, over several visits, has been investigated as a risk factor for HPV persistence. Women with a co-infection with two or more HPV types at the same visit were observed to be more likely to have a same-type persistent HPV infection at the subsequent visit, and the longer the infection persisted, the harder it was to clear [Ho et al., 1998]. However, the clinical significance of co-infections is still not clear in light of two recent studies that showed that acquisition of a new HPV infection was more likely among women with a prevalent HPV infection but that persistence of an HPV infection was independent of co-infection with other HPV types [Liaw et al., 1999; Rousseau et al., 2001].

Despite the varying degrees of association between HPV persistence and viral determinants, the characteristics of the viral infection do not fully explain why only a small proportion of women will experience a persistent infection, given that the vast majority of low-risk and HR-HPV infections appear to be transient [Ho et al., 1998].

The host immune response is thought to be of critical importance in the maintenance of an HPV infection. The HLA genes, particularly the class II HLA alleles, are the primary mediators of cell-mediated immune system responses to viruses and are highly polymorphic [Maciag & Villa, 1999]. They are expressed in antigen-presenting cells such as macrophages and Langerhans cells [Ferenczy & Franco, 2002] and play a major role in regulating T-cell response to foreign antigens. Inherited alleles may be a contributing factor in the outcome of HPV infections, although data on the precise role of HLA

haplotypes has been mixed. Many studies have observed an increased risk of HPV infection and cervical disease in individuals with DQB1*03, regardless of HPV type [Wank et al., 1992; Sanjeevi et al., 1996; Odunsi et al., 1996; Hildesheim et al., 1998; Cuzick et al., 2000] although at least one study was not able to demonstrate an association between HPV 16 related disease and any DQB1*03 alleles [Bontkes et al., 1998]. In contrast, DRB1*13 alleles appear to confer a protective effect against developing cervical disease [Apple et al., 1994; Ellis et al., 1995; Odunsi & Ganesan, 1997].

The inactivation of the *p53* tumour suppressor gene, in the human genome, can also facilitate viral integration [Ferenczy & Franco, 2002] and a recent study observed that a *p53* polymorphism resulting in the replacement of a proline with an arginine on codon 72 was more susceptible to HPV-E6-mediated degradation [van Duin et al., 2000]. Since then, a few studies with access to biopsies, have observed an increased risk for cervical cancer among women with the *p53* Arg/Arg genotype compared to women with the *p53* Pro/Arg or Pro/Pro genotype at codon 72 [Zehbe et al., 1999; Makni et al., 2000; Agorastos et al., 2000]. However, this finding has been refuted by many other studies [Joseffson et al., 1998] [Rosenthal et al., 1998; Klaes et al., 1999; Bertorelle et al., 1999]. Methodological differences between studies, including inter-laboratory variation, may partially explain the discrepancy [Makni et al., 2000]. Furthermore, the majority of these studies were based on the availability of stored biopsies that were convenient to test, but lacked an appropriate comparison group that was representative of the source population from where the cases had occurred. This convenient study design may have introduced selection bias into many of these studies, particularly if the comparison group differed from the target population with respect to both the exposure and outcome, resulting in a distorted measure of effect.

2.3.4.2 Other factors

Other factors that have historically been related to cervical cancer including hormonal factors (use of oral contraceptives and multiparity), other STDs (HSV-2, *Chlamydia trachomatis*), tobacco use, and dietary factors, are now being re-evaluated as potential predictors of HPV persistence and/or co-factors in cervical carcinogenesis [Bosch et al.,

1992; Schiffman et al., 1993; Evander et al., 1995; Kjaer et al., 1996; Deacon et al., 2000]. Some of these variables might have endogenous or exogenous hormonal influences on the attenuation of the immune response to the virus, thus, facilitating propagation and persistence.

The majority of HPV natural history studies, among women without cytological abnormalities at enrollment, have observed a positive association with age and HPV persistence. In general, older women (>30) are more likely to have a persistent infection [Hildesheim et al., 1994; Ho et al., 1998] [Ahdieh et al., 2001] compared to younger women. However, at least one study observed an inverse relationship with age [Brisson et al., 1996]. The classic markers of sexual activity, including age at first intercourse, have not been identified as predictors of HPV persistence in most studies, although one study observed that persistence was inversely associated with number of lifetime partners and positively associated with number of recent sexual partners [Hildesheim et al., 1994].

Estrogen may reduce susceptibility to primary HPV infection but in the event of a persistent HPV infection, sex hormones (estrogen and/or progesterone) may be associated with progression to cervical cancer [Brabin, 2002]. Hormonal binding can affect the regulatory functions in the viral gene expression [zur Hausen, 1989] and it has been suggested that OCs might promote the carcinogenic properties of HPV [LaVecchia et al., 1996]. However, the evidence for the use of oral contraceptives (OC) and persistent HPV infections is not convincing [Bosch et al., 2002]. The majority of cohort studies have not observed an association with duration of oral contraceptive use and HPV persistence [Hildesheim et al., 1994; Moscicki et al., 1998; Ho et al., 1998], although at least one study observed that women who used hormonal contraceptives for more than 6 years were nearly 4 times (OR=3.6, 95% CI: 1.2, 10.5) more likely than non-OC users to have a persistent HPV infection [Brisson et al., 1996]. However, persistence was only based on positivity at 2 visits, (over a 10 week interval, on average), in this last study. Evidence from the IARC multicentre case-control study provides strong evidence that oral contraceptives may act as a late-stage co-factor in HPV induced carcinogenesis [Moreno et al., 2002]. However, whether or not sex steroids including hormonal contraceptives

enhance the propagation or persistence of the virus, before the development of precancerous lesions, still needs to be further investigated.

Tobacco use has been postulated to interfere with the host immune response thereby facilitating a persistent HPV infection [Palefsky & Holly, 1995]. However, while an increasing number of studies have observed an association between current smoking and cervical neoplasia, after controlling for HPV exposure [Kjaer, 1998; Schiff et al., 2000; Deacon et al., 2000; Kjellberg et al., 2000; Munoz et al., 2000; Moscicki et al., 2001; Hildesheim et al., 2001a] the opposite has been observed for HPV persistence. Two prospective cohort studies surprisingly reported that moderate smoking was protective against a persistent infection with the same HPV type [Hildesheim et al., 1994; Ho et al., 1998]. Nonetheless, one other cohort study recently observed that HR-HPV infections amongst women who smoked persisted for 2 months longer than nonsmokers and a dose-response with duration was observed by the authors [Giuliano et al., 2002]. This last study had the highest one-year cumulative incidence for HPV infections in the literature and may represent a higher risk group compared to some of the young university cohorts. Consequently, Giuliano et al [Giuliano et al., 2002] may be able to capture the effects of heavy smoking on HPV clearance with more power than previously described in the literature. Nonetheless, tobacco use needs to be further evaluated to determine if the direct effect of the tobacco carcinogens facilitate HPV acquisition and persistence or if they occur further downstream and act as a cofactor with a promoter effect on an established HPV infection, thereby influencing HPV related neoplasia, but not necessarily HPV persistence [Ho et al., 1998].

Two seemingly opposing mechanisms for tobacco exposure emerge that suggest the early stages of HPV infection and colonization of the epithelium is not necessarily facilitated by tobacco-induced immunosuppression, and tobacco exposure may even hinder HPV persistence. It is also possible that moderate smoking may be correlated with other unmeasured lifestyle habits that may protect a woman from a persistent HPV infection [Hildesheim et al., 1994]. Despite the detection of tobacco metabolites in the cervical mucus in smokers, the direct effect of the tobacco carcinogens may only occur further

downstream and act as a cofactor with a promoter effect on an established HPV infection, thereby influencing HPV related neoplasia, but not necessarily HPV persistence [Ho et al., 1998].

It is not clear whether HIV immunosuppression worsens the natural course of HPV infections [Clarke & Chetty, 2002], permits reactivation of previously latent HPV infections, or facilitates the rapid infection of HPV [Schiffman & Brinton, 1995]. Recent data reviewed by Clarke and Chetty [Clarke & Chetty, 2002] suggests that the relationship between HIV, HPV and cervical cancer is quite complex, in light of results from researchers Sun et al. [Sun et al., 1995] who were unable to demonstrate a direct correlation between oncogenic HPV persistence and CIN against a varying degree of immune suppression, among a group of HIV seropostive women and controls. The detection of HIV in macrophages has also led some researchers to speculate a potential mechanism for HPV-HIV interaction because the local cervical immune response is a crucial factor in HPV replication [Clarke & Chetty, 2002].

Chlamydia trachomatis has been associated with cervical disease [Schiff et al., 2000; Smith et al., 2002; Bosch et al., 2002], possibly by acting as a cofactor in HPV-induced tumorigenesis via an inflammatory pathway [Castle et al., 2001]. In contrast, studies that have evaluated an earlier endpoint of cervical disease (SIL) or HPV persistence have generally failed to observe an association with *Chlamydia trachomatis* [Ho et al., 1998; Moscicki et al., 2001; Sellors et al., 2003] [Hildesheim et al., 2001a].

Little is known about the effect of alcohol and HPV persistence. It has been hypothesized that women who consume excessive amounts of alcohol (> 2 drinks per day) may possess higher circulating levels of estrogen and that the vaginal epithelium, including the cervix, may be more estrogen responsive [Hankinson et al., 1995] [Reichman et al., 1993]. In a recent population-based cohort study in Sweden, women discharged from hospital with a diagnosis of alcoholism between 1965 and 1995 were followed and linked to national cancer registries [Weiderpass et al., 2001]. The cohort was nearly twice as likely to have *in situ* carcinoma, and had a three-fold increase risk for cervical cancer, (standardized

incidence ratios were computed based on nationwide specific cancer rates). However, HPV status was not measured, and although information on sexual activity was not available, it is reasonable to hypothesize that this specific population might be more likely to engage in high-risk sexual activities (e.g. multiple sex partners, infrequent Pap smear screening and unprotected sex) which could very likely explain the observed effect, rather than alcohol abuse. Nonetheless, increased alcohol consumption was associated with acquisition and persistence of HPV in one cohort study of young female students [Ho et al., 1998], but not another [Moscicki et al., 2001].

Diet may play a role in the steps following acquisition of an HPV infection. Low folate levels in the diet has been thought to be a risk factor for cervical carcinogenesis because folate is required for DNA synthesis, repair and methylation [Eto & Krumdieck, 1986; Butterworth et al., 1992; Mason & Levesque, 1996]. Low folate levels may also facilitate the incorporation of HPV into the host genome, since the virus has been observed to integrate into the host DNA of several cervical cancer cell lines at fragile sites made susceptible to breakage by inadequate levels of folate [Popescu et al., 1987; Gallego et al., 1994]. Regression of cervical lesions was observed in one clinical trial with folic acid supplements [Butterworth et al., 1982] but not in two others [Butterworth et al., 1992; Childers et al., 1995].

A number of other dietary or circulating micronutrients have been proposed as protective agents against cervical neoplasia. Vitamin A (retinol) and the carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, or lycopene) are important dietary components that are thought to counter the action of some carcinogens, either by reducing their electrophilic potential (anti-oxidants) or by inducing normal maturation and differentiation of the cervical epithelium [Franco, 1993; Sedjo et al., 2002a]. Two early studies that measured circulating nutrient levels in the serum had contradictory results. Giuliano et al. [Giuliano et al., 1997] observed lower serum levels of β-carotene, β-cryptoxanthin, and lutein with HPV persistence, while Palan et al. [Palan et al., 1998] did not observe an association of HPV persistence with mean levels of circulating retinal, α-carotene, β-carotene or lycopene when compared to women without an HPV infection or with a transient

infection. In a more recent study assessing both dietary and circulating levels of micronutrients in a cohort of young women from Arizona, a higher level of vegetable consumption was associated with a 54% decrease risk of HPV persistence [Sedjo et al., 2002b]. Persistence was defined as a woman having any oncogenic HPV infection at two or more consecutive time points and women who were negative throughout the study were excluded from the analyses. The study also observed that women with the highest level of plasma *cis*-lycopene or circulating vitamin B_{12} were less likely to have a persistent infection compared to women in the lowest tertiles [Sedjo et al., 2002a]. A dose response was also observed between increasing vegetable fibre in the diet and reduced risk of HSIL, after adjusting for HPV status, smoking and age, in a population-based case-control study in Sweden, [Kjellberg et al., 2000].

2.4 Summary of literature

In summary, cervical cancer is the leading female malignancy in most developing countries. In developed countries with good screening programs, invasive cervical cancer is relatively rare, whereas its precursors and equivocal cytological results represent a major health burden. HPV is present in 5-40% of the world's female population, at any time, and it has been shown that, under certain conditions, these sexually transmitted infections can persist and induce precancerous cervical lesions and cervical cancer [Bosch, de Sanjosé 2003].

There is a tremendous effort to introduce new preventive strategies, such as prophylactic vaccines against HPV, into high-risk populations. However, important issues need to be resolved before the introduction of HPV vaccination in different populations, such as the determination of how many HPV types, other than HPV 16, need to be targeted in the vaccine. Furthermore, once approved by federal health agencies, HPV vaccines will be of most benefit to young adults who have not yet been sexually exposed to the virus. It will be decades before all birth cohorts can benefit from prophylactic vaccination against HPV.

Thus, epidemiologists are still very interested in understanding what lies between the causal exposure and the disease endpoint, namely, the natural history of HPV leading to genital neoplasia. High parity, smoking, and long-term use of oral contraceptives have been established as cofactors for cervical cancer among women with persistent infections with HPV. Other sexually transmitted diseases such as Chlamydia trachomatis, HSV-2 and HIV as well as some poorly known dietary factors may be intervening factors. The future focus of natural history projects will likely concentrate on defining the risk factors and biomarkers for HPV clearance versus persistence and progression to precancer [Schiffman, Kruger Kjaer 2003].

2.5 Rationale

Given that HPV testing is being considered as an adjunct to Pap smears in future cervical cancer screening programs, there is a need to better understand the epidemiology of HPV infections. Characterizing the burden of HPV infections in different populations and estimating the average duration of these viral infections will help in the development of a clinically relevant definition of HPV persistence. Furthermore, if there are environmental factors that can influence the rate of HPV clearance, then data from natural history studies on HPV are of paramount importance, given the current paucity of data on potentially modifiable co-factors for HPVpersistence.

CHAPTER 3: OBJECTIVES

The objectives of this study are to:

1. Describe the incidence and clearance rates of overall, high oncogenic-risk, low oncogenic-risk and type-specific HPV infections (Manuscript I)

2. Identify determinants of HPV acquisition (Manuscript II)

3. Identify determinants for HPV clearance (Manuscript III)

4. Identify viral determinants of low-grade squamous intraepithelial lesions (Manuscript IV)

CHAPTER 4: MATERIALS AND METHODS

4.1 Study design

4.1.1 Subjects

Female students attending either the McGill University Health Clinic or the Concordia University Health Clinic were invited to participate in this prospective cohort study. Women were eligible if they intended to be in Montreal for the next two years and had not had an abnormal Pap smear in the last 12 months or required treatment for cervical disease in the last year. Recruitment was initiated in November 1996 at McGill University and in November 1997 at Concordia University. Subject accrual was completed in January 1999 and follow-up was terminated on November 30, 2001. Testing for additional host and viral biomarkers has proceeded as part of a renewal grant from the Canadian Institutes of Health Research but its results are not included in this thesis.

A nurse practitioner at each site was responsible for introducing the study to all females waiting to see a nurse or physician in the reception area. The study was also advertised in the university newspapers, radio (McGill-CKUT), and in person (H. Richardson) to first year classes at McGill. At baseline, participants were asked to sign a consent form and complete a personal data sheet and questionnaire (Appendices I.I – I.III). Pap cytology and cervical cell specimens were obtained after the consent form was signed. The women enrolled in the study were asked to return to the clinic every 6 months over a period of 2 years, for a total of 5 visits. At each of the return visits participants completed a modified (follow-up) questionnaire (Appendix I.IV) designed to measure recent behavioural changes and had cervical specimens taken for Pap cytology and HPV testing.

4.1.2 Study Population

There were six hundred and thirty five women who originally consented to participate in the study. However, fourteen women were withdrawn from the study because they did not return for the second visit and, either had a sample that could not be analyzed (7%) or did not complete the baseline questionnaire (93%). The study population thus consisted of 621 subjects with 2688 visits.

Because the study had many different outcomes (acquisition of a HR-or LR-HPV infection, clearance of a HR-or LR-HPV infection, regression of a prevalent or incident squamous intraepithelial lesion), if a woman was not at risk for the outcome of interest she was not included in the analyses for that specific outcome. The number of women included in each analysis is described in more detail in the section on statistical analysis.

4.1.3 Participation rate and compliance

A cross-sectional epidemiologic study of over 500 female students recruited between 1992-1994 was conducted at McGill university and estimates of subject participation, HPV prevalence, rate of cytologic abnormalities, and of the distribution of sociodemographic variables and other risk determinants in the population were obtained[Richardson et al., 2000]. Results from the pilot study suggested that the McGill Health clinic performs, on average, 1500 annual consultations in which Pap smears are taken. Approximately one-fifth of these consultations represent follow-up examinations for abnormal findings and would thus be ineligible. Of the approximately 1200 initially eligible annual routine Pap smear consultations, an estimated one-third were expected to be ineligible because of the students' inability to comply with the 2-year follow-up schedule or because of refusal to participate, leaving 800 women per year potentially eligible for the follow-up study.

While the response rate for agreeing to participate in the cross-sectional survey was approximately 80%, it became evident that we could not obtain a similar level of voluntary participation, within one year, for a study requiring multiple returns to the clinic over a 2-year period. Of the 1700 women that had a Pap smear test at McGill University, in the first year of recruitment, 284 women agreed to participate (16.7%) in our cohort study. Consequently, in November 1997 the target population was expanded to include female students attending the health clinic at Concordia University. A similar proportion of the 1285 women having Pap tests at Concordia University (15.6%) agreed to participate during the 1997/98 autumn and winter terms. Therefore, our sample population represents about one-sixth of the target population. However, if we consider the assumption that only a proportion of the women at the McGill and Concordia health

clinics were eligible for the follow-up study, as previously discussed, then the sample population may represent as much as one-third of the potentially eligible target population.

Table 4.1 presents the number of visits and the follow-up rate per year in the cohort. Because of the importance of retaining all subjects for the entire duration of the study, participants were remunerated $20 for every return visit that they completed. The follow-up rate was calculated by comparing the number of subjects who returned for a particular visit, prior to November 30, 2001 to the number of subjects eligible for that follow-up visit. Women returned, on average, every 7 months for their follow-up visits and of the 621 women in the original cohort, (based on those women who had completed the first questionnaire), 424 (68%) completed all 5 visits.

Table 4.1
Time interval between visits and follow-up rates for the cohort of 621 women

Interval between visits	Mean number of months between visits (median)	SD	Range	Number of subjects returning for subsequent visit	Follow-up rate*
1 and 2	6.86 (6.28)	2.72	2.0-25.8	578	93.0%
2 and 3	6.57 (5.92)	2.81	1.0-26.2	553	95.7%
3 and 4	6.76 (6.02)	3.06	2.1-29.4	498	90.0%
4 and 5	6.66 (6.15)	2.87	1.0-21.7	424	85.0%

**The estimation of each follow-up rate was conditional on a woman completing the previous visit. There were no constraints on time between visits in this calculation.*

4.1.4 Data management

The data management was conducted at McGill University, Department of Oncology, Division of Cancer Epidemiology. Students who agreed to participate were given a numeric identification code (study id) that did not change throughout the follow-up. Each questionnaire was identified with the study id and a suffix (1 to 5) that indicated the visit

at which the questionnaire was completed. The cervical samples were identified with the same study id and suffix as the questionnaires.

The study nurse at each site was responsible for updating the respective personal database for McGill and Concordia. The personal database was used to log the number and date of each visit that a participant had completed and served to remind the study nurses when a participant was due for her next visit. Participants who were overdue for their visit (>8 months) were pursued more aggressively; additional reminder calls were made, and new addresses and phone numbers were identified, in the case when participants had moved.

The questionnaire data from each visit was stored in a separate database (maintained with the database management software Filemaker Pro 4) specific to the baseline or follow-up visit. A total of 5 questionnaire databases were created. The questionnaire data was entered by a student working in the Division of Cancer Epidemiology during the summers of 1998-2000 and by the author (HR). The McGill study nurse and the summer student verified that the data entry was accurate, for the baseline questionnaire, by comparing the answers from the original questionnaire with the data entered in the database. Data from the follow-up questionnaires were verified at the same time that the data were cleaned and recoded. Any variable with suspiciously high or low values was cross-referenced with the answers from the original follow-up questionnaire.

The data with the HPV results were received from the laboratory twice a year. There were data on variables corresponding to 27 HPV types and β-globin status. Once a year, a list of study id's, corresponding to each participant who had returned for visit x, (according to records in the questionnaire database and the personal database for visit x) and their respective laboratory result and cytology result at visit x was generated. If there were any instances that a woman had completed a visit but did not have an available HPV result, the laboratory was notified and asked to submit the result as soon as possible. If the cytology result had not yet been recorded in the study database a copy of the cytology report from the participant's medical chart was obtained from the clinic and the database was appropriately updated.

4.1.5 Generalizability

The target population in this study was female students attending the two aforementioned universities in Montreal, who utilize the student health clinics. Nonetheless, eligibility requirements restricted the target population to women in the first or second year of her studies so that she could commit to returning to the clinic over a two-year period. Subject selection was partly dependent upon self-selection and partly dependent on a proportion of the physicians from the two clinics that were willing to participate and help accrue subjects. The former may be a source of potential seletion bias so that discrepancies between risky behaviours of subjects included in our cohort and non-participating female students may exist. It is possible that women who volunteered to participate in our study may have been more health-conscious and consequently engage themselves in less risky behaviour than those women who refused to participate, which might result in a lower prevalence of risk factors and possibly infections. However, this would not necessarily bias our associations. In conclusion, the study population may only be representative of female students who use the university health clinics and voluntarily participate in educational health research programs. Of the 2995 women who had a Pap smear at McGill (1709) and Concordia (1285), in the period of one year, 16% of the women (484/2995) were accrued into our cohort (284/1709 at McGill and 200/1285 at Concordia). Therefore, our sample population represents approximately one-sixth of the target population.

Nonetheless, an effort was made to compare certain outcomes of the cohort with outcomes of the non-participants at the two clinics. The overall number of cytology tests and respective results for all female students attending the clinic, in a one-year period, were made available to us. This data was anonymous but allowed us to compare the proportion of women in our cohort who had abnormal cytology in a year with the corresponding target population. Women in the target population (attendants of the university health clinic) experienced very similar occurrences of cervical lesions to the women in our cohort, within a one-year time span; 97.7% had normal cytology, 2.2% had an LSIL diagnosis and 0.1% had a diagnosis of HSIL.

In addition, information on certain personal behaviours and characteristics that were measured from the Canadian Campus Survey (CCS) in 2000 [CAMH, 2000] and the National Population Health Survey (NPHS) in 1996/97 [Statistics Canada, 2003], restricted to college and university students, were reviewed and the distribution of relevant factors were compared to our cohort. Table 4.2 describes the distribution of available exposures among our cohort with the populations sampled in the two Canadian surveys.

The women from the NPHS survey that were eligible for inclusion as a comparison group, were those women who were currently in school, were between 20-45, and had already received a post-secondary degree. The women in this NPHS sample were older, on average than the women in our cohort. Only 47.6% of the women were less than 25 years of age, compared to 73.8% of the women in our cohort. In terms of risk behaviour, the distributions of certain markers of sexual activity were comparable. Approximately half of the women had initiated sex before the age of 18 in both the NPHS sample and our cohort, and 95% of the women did not have an STD infection in the last year. More women used oral contraceptives in our cohort compared to the group in the NPHS survey. There were slightly more women who always used condoms, in the NPHS survey, but condom use was comparable between those women in the Campus survey and our cohort.

Women in our cohort were slightly less likely to be current or former smokers compared to women in the NPHS group but there were more daily smokers in our cohort (25.0%) than in the NPHS group (19.4%) or in the Canadian Campus survey (12.1%), although the survey did not provide an estimate for female students only.The average number of alcoholic drinks per week reported in the Canadian Campus survey was 3.9 drinks/week among female students, which was slightly higher than the average number reported in our study or the NPHS survey. Thus, while there were some differences between the groups of women, most were not substantial, and it is reasonable to suggest that the study findings can have some application for other female students in Canada.

Table 4.2

Distribution of selected exposure variables for the Montreal University Student Cohort, post-secondary students in the NPHS (1996/97) survey and the Canadian Campus Survey (2000)

Variable	Montreal Cohort* N =621	NPHS 1996/97 N=2224	Canadian Campus Survey N=7800
Age at first sexual encounter			
12-14 years	8.1%	6.3%	N.A.
15-17 years	46.0%	40.6%	
18-19 years	34.0%	32.4%	
20-24 years	11.6%	17.2%	
25-29 years	0.5%	3.1%	
Number of sex partners in last year			
0-1	87.5%	61.9%	N.A.
2	7.5%	19.8%	
3+	5.0%	18.3%	
Oral contraceptive use:			
Never in last year	47.2%	71.7%	N.A.
Sometimes in last year	4.8%	N.A.	
Always in last year	48.0%	28.2%	
Condom use:			
Never in last year	26.9%	15.0%	27.9%
Sometimes in last year	27.5%	30.4%	N.A.
Always in last year	45.6%	54.6%	28.9%
History of *STDs***			
Never in last year	94.5%	95.7%	N.A.
Ever in last year	5.5%	4.3%	
Smoking status			
Never	60.2%	50.2%	N.A.
Former	15.8%	27.0%	
Current	24.0%	22.8%	
Proportion that consumed cigarettes daily	25.0%	19.4%	12.1%
Mean # alcoholic drinks consumed weekly	3.1	3.3	3.9
SIL status at visit 1:			
Normal	97.1%	N.A.	97.3%
LSIL	2.7%		2.7%
HSIL	0.2%		0.2%

**Variables measured at baseline,*
***STD status does not include genital warts,*
**** N.A.= Data not available*

4.2 Measurement

4.2.1 Risk behaviour

Information on risk behaviour was collected at enrollment from a self-administered questionnaire that had been used in a previous cross-sectional study among the same target population [Richardson et al., 2000]. The measurement instrument was a compilation of questions developed and validated by the National Cancer Institute and the International Agency for Research on Cancer (IARC), focusing on the measurement of different dimensions of sexual behaviour and hygiene, and use of contraceptives, tobacco and alcohol. Nonetheless, it should be noted that, given the exploratory nature of this study, especially with regards to examining predictors of HPV clearance, measurement of some of the variables including dietary habits was fairly crude. However, at the start of this study, there was virtually no information about risk factors for HPV persistence and the intention of this study was to provide general insight into some potential predictors of HPV clearance that could be further investigated in future studies.

While the questionnaire that was developed for this study was never validated, data on the repeatability of measuring sexual activity and reporting error was compared with results from this study in Montreal and five other cohort studies based in Denmark, Costa Rica, San Francisco, Toronto and Sao Paolo. The strongest predictors of reporting error of sentinel questions such as "age at first intercourse" and "number of lifetime partners", reported at multiple repeat visits, included age, ethnicity, education and cohort membership [Schlecht et al. 2001]. The lowest discepant rates for any reporting error was observed in the Montreal (McGill/Concordia University) cohort, and the between interview agreement was >90%. These results suggest that our cohort members were able to understand the questions in our questionnaire and provide reliable answers at multiple repeat visits.

The questionnaires were in English and French and in addition to the aforementioned risk factors in the previous passage, the questionnaire also inquired about socio-demographic status, race, diet, reproductive history and medical history (Appendix I.III). An abridged (follow-up) questionnaire (Appendix I.IV) designed to measure changes in recent sexual

practices and other lifestyle factors, was completed at each subsequent visit. Participants were given the opportunity to answer the self-administered questionnaires in a private office at the health clinic after their physical examination. The study nurse was available at all times to answer any questions they might have about the questionnaire or any other health related matter. Once they completed the questionnaire they were given $20 and asked to sign a waiver confirming that they had received their remuneration.

4.2.2 Variable Description

The risk factors that were individually evaluated in this project were selected from sets of behaviours or activities measured in the baseline and follow-up questionnaires, based primarily on putative risk factors for HPV acquisition or clearance, described in the literature. The baseline questionnaire had eight sections including:

1. General information on age, ethnicity and parents' occupation;
2. Diet history on foods rich in vitamins A, C, D, E and B-complexes;
3. Smoking history and alcohol consumption;
4. Reproductive history;
5. Sexual history including number of oral and vaginal sex partners, frequency of oral and vaginal sex, and other intimate sexual practices including anal sex and masturbation;
6. Contraceptive history on the duration of oral contraceptive use and barrier methods of contraceptives;
7. Personal hygiene habits including ablutionary practices such as daily bathing and douching;
8. Medical history including frequency of Pap smear screening, history of sexually transmitted diseases and use of hormone therapy.

Table 1.i in Appendix I.V summarizes the risk behaviours and characteristics that were investigated in this project and describes the distribution of the selected variables at enrollment. In the case of variables that were originally continuous, the mean, median and standard deviation for the continuous variable are presented alongside the frequency of the newly created categories. The frequency of missing values for each variable at

enrollment, are presented in the left column of table 1.i with the variable description. The right column of table 1.i represents the recoded variables, including the recoded missing variables. The strategy used in this study for managing missing data is addressed in a later section of this chapter.

Age at enrollment was calculated by subtracting the date of enrollment into the study from the participant's date of birth, and was categorized into four levels, based on observations that the rate of HPV infection declines with age particularly after age 30 for HPV incidence [Ho et al., 1998], HPV persistence and cervical neoplasia [Hildesheim et al., 1994]. Race was based on the participant's country of birth and ethnic representation. Based on previous studies that have observed increased risk for HPV and cervical disease in women of Black and/or Hispanic descent, [Wheeler et al., 1993; Ho et al., 1998; Villa et al., 2000] four categories were created. The majority of women who classified themselves as French or English Canadian, Jewish Canadian, American, European, Australian, or from New Zealand or the Middle East were defined as "White". However, if they also specified Asian, Hispanic or African/Caribbean descent, they were classified as Asian, Hispanic or Black, respectively.

The diet section was composed of a few food frequency questions that attempted to capture the participant's general intake level of some important nutrients required for cell functioning and repair, such as calcium (found in dairy products) and folate (present in dark leafy greens) and anti-oxidants such as vitamin C & E and lycopenes (such as found in tomatoes) that prevent cellular damage from free radicals. Questions were asked about the participant's consumption patterns for milk, cheese, yogurt, common fruits and vegetables, citrus juices and vitamin C-supplements and foods rich in iron, such as liver. Participant's could respond that they consumed the given product 1+/day, <1/day but >1/week, 1/week, 1/month or never. The consumption of dairy products and vegetable products were grouped into separate aggregate variables based on composite measures and dichotomized, as described in table 1.i.

Some of the variables only refer to information collected at baseline and are thus considered as time-fixed variables. Nonetheless, the majority of variables refer to recent behaviours or activities that were collected at each subsequent visit, and are therefore, considered as time-dependent variables. All the time-dependent covariates captured information based on behaviour since the last visit. However, at baseline, the corresponding question referred to information based on behaviour throughout the participant's adult lifetime.

Several variables had a cumulative component to them such as tobacco use and duration of OC use. Pack-years was based on a calculation that combined intensity and duration of tobacco use, and the values would change with each visit, if the participants were smokers or had just initiated smoking. If a participant was no longer consuming tobacco since her last visit, then her value for pack-years at the previous visit was carried forward. A similar strategy was used for calculating overall duration of OC use.

The assumptions for linearity were assessed for variables that had continuous values (see Statistical Analyses for more details). The only variable that appeared to violate the assumptions of linearity was number of lifetime partners, which was categorized into 4 levels. It was also decided to categorize the other continuous variables for ease of comparison with other results in the literature. Categories were chosen for continuous variables that represented meaningful partitions with some consideration given to the need to balance the size of the categories (table 1.i). The creation of categories may have introduced some misclassification of exposure. However, it also minimized the impact of extreme values in the data analysis, and made it easier to evaluate dose-response relationships.

4.3 Laboratory methods

4.3.1 Sample collection

To detect human papillomavirus infection with maximum sensitivity, cells must be collected and assayed for HPV DNA. Cervicovaginal lavage and cervical cell scrapes are two methods employed for sampling exfoliated cervical cells. Cervicovaginal lavage appears to be the more sensitive collection method for exfoliated cervicovaginal cells for the detection of HPV. However, the lavage method might also collect vaginal cells. Therefore, epidemiological studies investigating cervical HPV infections generally use the more conservative cervical cell scrape method to minimize misclassification from the possible detection of vaginal HPV infections [Goldberg et al., 1989; Morrison et al., 1992]. The doctors at the university health clinics employed the scrape technique using two Accelon combi cervical biosamplers (Medscand Inc., Hollywood, Fla.) which samples both the ectocervix and endocervical canal, simultaneously. The first sampler was used primarily for a Pap smear with the remaining cells and the cervical cells collected from the second sampler used for HPV DNA testing. The plastic sampler tips containing the exfoliated cells were placed in a tube with 2 ml of saline-buffer. The tube was briefly agitated, the samplers were then discarded and the tube containing the cell suspensions was frozen until delivery to Dr. Coutlée at the Laboratoire de Virologie Moleculaire, Centre de Recherche et Département de Microbiologie et Infectiologie, Hôpital Notre-Dame du Centre Hospitalier de l'Université de Montréal (CHUM).

4.3.2 Clinical samples

At the Microbiology laboratory (CHUM), the collected cells were briefly agitated (<15 seconds) in a tube containing 2 ml of 10mM Tris-HCl [pH 7.4] and 0.1 mM EDTA (TE). For the first 354 samples, 200 µl of the cell suspension was lysed with Tween 20 (final concentration, 0.8% [vol/vol]), digested with 250 µg/ml of proteinase K at 45°C for 2 hours, purified with GlassMAX resin (Gibco-BRL, Burlington, On. Canada) and then resuspended in 50 µl of TE saline-buffer. For the remaining samples (n=2334), the cell suspensions were processed with QUIAamp columns (QUIAGEN Inc., CA. USA), a system that purified DNA more quickly and efficiently. Purified DNA was boiled for 10 minutes and then stored at -70°C[Coutlée et al., 1999]. The remaining (unlysed)

specimens were stored at -80°C until needed.

4.3.3 HPV DNA detection

4.3.3.1 ß-globin and HPV amplification

A polymerase chain reaction (PCR) system was used to detect human β-globin and HPV DNA. The β-globin detection system was used as an internal control to demonstrate the absence of inhibitors, the presence of an adequate number of epithelial cells and the integrity of the processed DNA [Bauer et al., 1991; Bauer et al., 1992]. Coamplification of β-globin and HPV DNA was avoided because of the competition between both target amplifications as demonstrated by our collaborators in their laboratory [Coutlee et al., 2002]. Human β-globin was detected in 5 µl of purified DNA with specific primers, GH20 and PC04, which target a 268 base pair (bp) region of the gene [Altman et al., 1995]. If human β-globin could not be detected, samples were extracted with phenol-chloroform and precipitated with ethanol. Five hundred nanograms of the extracted sample were then amplified for ß-globin. [Mayrand et al., 2000] If β-globin could not be detected in the extracted samples, the DNA sample was considered inadequate for further HPV testing.

β-globin-positive specimens were amplified separately with consensus primers MY09/MY11 and HMB01. These primers target the highly conserved late region of HPV that encodes the viral capsid L1 protein, amplifying a 450 bp fragment. The HPV detection system is a highly sensitive DNA amplification method that can detect as few as 1-10 molecules of HPV-16 from a genital sample. The system uses a mixture of degenerate primers, MY09 and MY11, to amplify a broad spectrum of HPV types [Manos et al., 1989; Bauer et al., 1991]. A third primer, analogous to MY09, HMB01, is added to enable detection of HPV 51 [Hildesheim et al., 1994]. The amplification mixture contained 6 mM $MgCl_2$, 50 mM KCl, 7.5 U of *Amplitaq Gold* DNA polymerase (Roche Molecular Diagnostics, Mississauga, On. Canada), 600 µM dUTP and 200 µM of each dATP, dCTP and dGTP and 50 pmol of each biotin-labeled primer (MY09, MY11, HMB01). The PCR profile for DNA amplification was used in a TC 9600 thermal cycler:

activation of AmpliTaq Gold at 95°C for 9 minutes; 95°C denaturation for 1 minute, annealing at 55°C for 1 minute, extension at 72°C for 1 minute for 40 cycles; 5 minute terminal extension at 72°C [Coutlée et al., 1999].

4.3.3.2 HPV genotyping

A reverse line blot method was used to identify 27 HPV types in a reaction in which biotin-labeled HPV amplicons (PCR products) were hybridized to an array of immobilized oligonucleotides, for 27 genotypes, on a single "strip". [Gravitt et al., 1998; Coutlée et al., 1999] Twenty-five microliters of Amplicor denaturation solution was added to 50 µl of PCR-amplified products. Seventy microliters of the denatured PCR product was added to each well of an Amplicor typing tray that contained 3 ml of hybridization solution (4 x SSPE, 0.1% SDS) prewarmed to 53°C and a strip of HPV oligonucleotide probes. The probe mixes for the following 27 HPV genotypes were fixed on distinct lines on each strip: types 6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 57, 58, 59, 66, 68, 73, 82, 83, 84. The tray was incubated in a shaking water bath at 53°C for 30 minutes The hybridization solution was aspirated from each well and 3 ml of washing solution containing 1 x SSPE and 0.1% SDS, was added at room temperature and aspirated. Again 3 ml of washing solution was added to each well and incubated at 53°C for 15 minutes. The washing buffer was aspirated and 3 ml of Amplicor streptavidin-horseradish peroxidase conjugate was added. The tray was shaken gently for 30 minutes at room temperature. The conjugate was aspirated and 3 ml of washing buffer was added. Trays were shaken for 10 minutes on a platform shaker. This step was repeated once. After aspiration of the washing buffer, 3 ml of citrate buffer was added to each well and was aspirated. The substrate was prepared by mixing 0.01 % H_2O_2 and 0.1% ProClin in a 0.1 M citrate solution, with 0.1% tetramethylbenzidine in dimethylformamide. Three milliliters of substrate was added to each well. The trays were shaken at 70 rpm for 5 minutes at room temperature. The substrate was removed and the strips were rinsed with distilled water, stored in citrate buffer and read within 30 minutes [Coutlée et al., 1999].

The protocol yielding the most sensitive results with MY09/MY11/HMB01 was used in this study as established in previous validation studies. [Coutlée et al., 1999; Gravitt et al., 2000] Consensus primer pairs have been shown to reach different sensitivity levels

depending on the type analyzed and the MY primer system appears to have a lower analytical sensitivity for certain genital HPV types including 35, 52, and 56 when compared to the PGMY or GP+ primers [Harnish et al., 2000; Gravitt et al., 2000; Coutlee et al., 2002]. It appears that the only way to improve the sensitivity of HPV DNA detection is to utilize more than one primer system, [Harnish et al., 2000] even when using the most sensitive PGMY and SPF assays [van Doorn et al., 2002]. Nevertheless, the MY09/MY11 primer set has been extensively used in epidemiological studies and has been well validated. It was also not feasible to change our protocol half-way through the study, when the advantages of the PGMY system were published, since the first half of the test results would no longer be comparable to the second half of the test results.

4.3.4 HPV 16 and HPV 18 variant analysis

Clinical samples that were positive for HPV types 16 or 18 at two or more visits were analyzed for HPV16 and 18 variants. A PCR-sequencing method was used with specific primers for HPV 16 and 18 which targeted a region within the long control region (LCR) that contains a hypervariable genomic segment [Ho, 1991; Chan et al., 1992; Ho et al., 1993; Franco et al., 1994; Myers et al., 1995; Yamada et al., 1995; Villa et al., 2000]. HPV-16 and HPV-18 LCR primers were designed to flank nucleotide positions 7478-7841 [Chan et al., 1992] and 7585-7805 [Ong et al., 1993], that correspond to their respective transcriptional enhancer [Chong et al., 1990]. This area of the LCR allows variants to be classified correctly since there is strong linkage of nucleotide changes between different regions of the genome [Chan et al., 1992; Yamada et al., 1995; Wheeler et al., 1997].

Amplification reactions were performed with 5 µl of processed sample in a 100 µl reaction volume containing 10 mM Tris-HCl [pH 8.3], 50 mM KCl, 2.5 units of Expand High Fidelity PCR enzyme (Boehringer Mannheim, Laval, Qué.), 0.5 µM of each primer, 0.25 mM each dCTP, dTTP, dGTP and dATP. The concentration of $MgCl_2$ was adjusted to 1.5 mM. The Expand High Fidelity PCR system is a mixture of *Taq* DNA polymerase and *Bwo* DNA polymerase that has a low rate of mistaken nucleotide incorporation and

increases the fidelity of PCR prior to sequencing [Mayrand et al., 2000]. Amplifications were completed in a 9600 Thermal Cycler (Roche Diagnostic System, Mississauga, Ont.). The amplification profiles with the LCR reagents included an initial step at 94°C for 120 seconds, followed by 10 cycles at 94°C for 15 seconds, at 55°C for 30 seconds, and at 72°C for 45 seconds. Amplification was completed with 35 cycles at 94°C for 15 seconds, at 55°C for 30 seconds, and at 72°C for an initial 45 seconds plus a five second increase in length per cycle, followed by an extension step at 72°C for 7 minutes. When the generated signals were too weak, amplifications were carried out for 60 cycles under identical conditions as above. If the signals were still too weak for sequencing, lysates were amplified with 10 units of *AmpliTaq Gold* DNA polymerase (Perkin-Elmer Cetus, Montréal, Qué), using the following amplification profile: activation at 95°C for 3.5 minutes, followed by 60 cycles at 95°C for 30 seconds, and at 60°C for 60 seconds initially plus a 2 second increase per cycle, followed by 10 minutes at 72°C. In each set of PCR reactions, negative controls were included to ensure the absence of contamination.

PCR-amplified HPV DNA fragments were purified with the QIAquick PCR purification kit (Quiagen Inc., Mississauga, Ont) when a single band was visualized after gel electrophoresis or with the QIAquick gel extraction kit protocol (Quiagen Inc.) when several bands were visualized. Direct double-stranded PCR-sequencing was done since most amplified products contain the appropriate nucleotide at any given position and the derived sequence will be a true representation of the HPV DNA template [Kaye et al., 1996; Yamada et al., 1997]. Twenty nanograms of the purified PCR product was sequenced with the fluorescent cycle-sequencing method (BigDye terminator ready reaction kit, Perkin-Elmer) with 25 cycles at 96°C for 10 seconds, 50°C for 5 seconds, and 62°C for 4 minutes. Sequence analysis was performed with an ABI Prism 3100 Genetic Analyzer system. If a non-prototype variant or ambiguities were obtained, PCR-sequencing was repeated once [Ong et al., 1993; Yamada et al., 1997]. If results between these 2 sequencing reactions were discrepant, 3 additional PCR-sequencing reactions were done to investigate the presence of multiple variant infections or *Taq*-induced

errors. In contrast to variants that had mutations in non-random nucleotide positions, unconfirmed mutations were considered as PCR artifacts [Xi et al., 1995].

When a mixture of variants was suspected after direct sequencing, PCR products were cloned using the TOPO TA cloning kit (Invitrogen) [Mayrand et al., 2000], after purification with QIAquick gel extraction kit, with the pCR2.1 TOPO vector and competent *E. coli* TOP10 strain (Mayrand 2000). Ten clones containing the HPV LCR fragment were identified by digestion with restriction enzymes *Hind*III and *Xho*I. The plasmid DNA from the transformed clones was purified using the QIAprep Spin Miniprep system (Quiagen Inc.) according to the manufacturer instructions and then sequenced. HPV variant sequences were compared to sequences of known HPV variants or prototypes using the BLAST sequence analysis from the Genetic Computer Group (GCG) [Altschul et al., 1990], (http//hpv-web.lanl.gov[Myers et al., 1995] and GenBank). Isolates with a DNA sequence different from the prototype strain were classified as non-prototype variants. Sequences from unknown non-prototype variants were aligned using the Clustal Multiple Sequence Alignment 1.8 program [Worley & McLeod] to further classify variants into the appropriate lineage.

4.3.5 Precautionary laboratory measures

Given the importance of preventing inter-specimen and reagent contamination, several preventive measures were taken including the physical isolation of sample processing and reaction areas, use of autoclaved solutions, prealiquoted and premixed reagents, use of diluted positive controls, use of disposable gloves, and use of filtered pipettes and "splash-free" tubes to avoid aerosol contamination. In addition to the above measures, a test-based approach to prevent contamination of specimens with amplified DNA was adopted.

Cervical carcinoma cell line HeLa (which contains 40 copies of HPV-18 DNA per cell) was obtained from the American Type and Culture Collection (Rockville, Md, USA) and maintained in Eagle's minimum essential medium supplemented with 10% fetal calf serum. Negative, weak positive (ten HPV-18 DNA copies), and strong positive controls

(HPV types 6/11, 16, 31, 33, 35, 39, 45), were included in each amplification run to monitor contamination and overall endpoint sensitivity of each PCR run.

4.3.6 Cytology

The Pap smears were prepared, with the first accelon biosampler, onto a glass slide and fixed in 95% ethanol. The cytology slides were processed and read by Ms. Juliette Robitaille in the laboratory of Dr. Ferenczy, at the Jewish General Hospital. Cytopathology reports were based on the Bethesda system for cytological diagnosis [Solomon, 1989]. The cytology reports were photocopied and the original was kept in the student's medical file. The data from the photocopy was entered into a database, identified only with the study id, and then filed in a secured filing cabinet.

4.4 Statistical analysis

A variety of statistical methods were used to analyze the data from this study. They are discussed in brief, due to space restrictions, in the four manuscripts. This section will describe, in more detail, the assumptions and relevance of the various statistical methods employed for the following articles as well as additional analyses, presented in the appendices II and III.

4.4.1 Defining time zero for time-to-event analysis

Change in exposure status and/or outcome status were two features that could be captured in this cohort study. Therefore, time zero could have been, in principle, defined at time that risk began, such as time at first exposure, or at time of enrollment into the cohort. However, defining time zero as time of first exposure would not have been feasible, since the objective was to explore the association of more than one potential exposure (risk factor) and HPV acquisition or clearance and, it would not have been possible to identify a start date at which each risk behaviour was initiated. Furthermore, most exposures had already been initiated before entry into the study and were, thus, left censored. Therefore, time zero for the analyses focusing on risk factors for HPV acquisition was defined as the time of entry into the cohort, for those women who were free of HPV at enrollment. For women who were HPV positive at their initial visit, time zero was defined as time (visit) that a woman became free of HPV, and was therefore, once more at risk for acquisition of

a new HPV infection. Time zero for the analyses focusing on risk factors for HPV clearance was defined as the time of the earliest visit that an HPV infection was first detected, including the initial visit for women who were HPV positive at enrollment.

While the relevant determinants associated with HPV at time t were measured at time t, the actual time period of exposure to the majority of risk factors pertained to the last 6 months ("since your last visit") prior to acquisition of an HPV infection or clearance of an HPV infection.

4.4.2 Estimation of incidence rates

The monthly incidence rates presented in manuscript one were calculated using the approach provided in equation (1). The incidence rate is the number of disease onsets in a population divided by the sum of the time period of observation for all individuals in the study population [Kleinbaum, 1996]. With respect to this study, the numerator consisted of the number of subjects that acquired a new HPV infection during the study period (n) and the denominator was the sum of person-time (women-months) that each subject contributed throughout the study period (L). The 95% confidence intervals (CI) were calculated using the Poisson distribution (refer to equation 2).

(1) $ID = n/L$

(2) $95\%CI = ID = \pm 1.96\sigma$, $Var = \sigma^2 = n/L^2$

4.4.3 Time-to-event analysis

Survival or time-to-event analysis consists of a variety of statistical procedures for data analysis for which the outcome of interest is time until an event occurs (e.g., in the context of this project, time to the occurrence of an infection or time to clearance of an infection). However, inherent to data from cohort studies is the reality that not all subjects will have the event of interest by the end of the study [Kalbfleisch & Prentice, 1980]. In addition, some participants may drop out of the study (be lost to follow-up) or withdraw from the study before the closing date of the study and before developing the outcome. Subjects who remain disease free to the end of the study or who are lost to follow-up are

censored observations. The investigator knows only that their time-to-event is longer than the censoring time but the exact time and disease status is unknown. Survival analysis accounts for both censoring of outcomes and unequal lengths of follow-up time.

4.4.3.1 Kaplan-Meier

The product limit method or Kaplan-Meier (KM) technique [Kaplan & Meier, 1958] is a non-parametric approach for analysis of survival data that does not have any underlying assumptions about the distribution of follow-up time. The estimated probability of remaining event-free until time t (S(t)) is equal to the cumulative product of the probabilities of surviving through each successive interval. In our analyses, these intervals are defined by consecutive times at which an HPV infection is detected or cleared [Altman et al., 1995]. Subjects who are censored prior to a given event are not considered beyond their censored time. The general formula for a KM survival probability until failure time $t_{(j)}$ is given by equation 3. This formula gives the probability of surviving past the previous failure time $t_{(j-1)}$, multiplied by the conditional probability of surviving past time $t_{(j)}$, given survival to at least $t_{(j)}$.

(3) $S(t_{(j)}) = S(t_{(j-1)}) * \Pr(T > t_{(j)} \mid T \geq t_{(j)})$

The cumulative risk for the development of the disease (outcome) is equal to 1 minus the overall probability of survival (1-S(t)). Statistical differences in the survival distributions between groups can be determined using the log-rank test, when survival curves are being compared [Kalbfleisch & Prentice, 1980]. The log-rank tests the null hypothesis that the survival curves in the two groups are identical, and has an approximate chi-square distribution with one degree of freedom [Peto et al., 1976]. KM analyses and graphs were generated with SPSS® version 11.0.

4.4.3.2 Cox Proportional Hazards Regression

Cox proportional hazards (PH) regression [Cox, 1972] is a method designed to analyse censored survival data and is based on the hazard function. Equation 4 describes the hazard function h(t), which denotes the instantaneous risk for the event to occur immediately after time t, given that the individual has survived up to time t [Kleinbaum, 1996].

$$(4)\ h(t) = \lim_{\Delta t \to 0} \frac{Pr(t<T<t+\Delta t \mid T>t)}{\Delta t}$$

Where T is the time-to-event of interest.

The Cox PH model is usually written in terms of the hazard function and gives an expression for the hazard (i.e. for HPV acquisition or HPV clearance) at time t for an individual with a given specification of a set of explanatory variables [Kleinbaum, 1996], as described by equation (5).

$$(5)\ h(t,\mathbf{X}) = h_o(t) \times \exp \sum_{i=1}^{p} \beta_i x_i$$

Cox PH regression is a nonparametric model because it avoids any parametric assumptions about the functional form of the baseline hazard $h_o(t)$ [Holford, 2002] which represents the hazard for the hypothetical reference group with "0" values of all p covariates. The Cox model estimates the regression coefficients (β_i's), i.e. logarithm of the hazard ratio, by maximizing the partial likelihood of model (5) [Cox, 1972]. Its maximization is carried out by comparing the covariate pattern of the "case" (who has an event-to-interest at time t) to the covariate patterns of all the subjects in the corresponding risk set. The risk set includes all subjects who are still "at risk" at time t, i.e. who are event-free and have not been censored by that time, including subjects who later become cases

The results of the Cox regression are presented in terms of the hazard ratio (HR). The estimated HR is computed by exponentiating the regression coefficient, β_I, of an independent variable of interest and is interpreted as the relative hazard associated with a unit increase in the covariate X_i assuming all other variables in the model are held fixed. Inherent in the Cox model is the assumption that the estimated HR comparing any two specifications of covariates is constant over time [Cox, 1972]. The HR comparing subjects X_i^* and X_i (where i=1,...p) is presented in equation (6).

$$(6)\ HR = \frac{h_0(t) \times \exp \sum_{i=1}^{p} \beta_i X_i^*}{h_0(t) \times \exp \sum_{i=1}^{p} \beta_i X_i} = \exp \sum_{i=1}^{p} \beta_i (X_i^* - X_i)$$

The baseline hazard rate cancels out so that the estimated HR does not depend on time t. Proportional hazards regression was conducted with SAS® version 8, using the PHREG procedure.

4.4.3.3 Verification of the Proportional Hazards Assumption through flexible models of time-varying effects

To ensure the validity of the Cox PH regression, (for manuscripts two and three), it is important to verify the PH assumption of a constant-over-time HR between categories of covariates [Altman et al., 1995]. The proportional hazards assumption was tested based on a flexible generalization of the Cox proportional hazards model [Abrahamowicz et al., 1996]. The flexible model allows the hazard ratio for selected exposures to change over time, according to an arbitrary function, the shape of which is estimated from the data using a quadratic regression spline [Ramsay, 1988] with 5 degrees of freedom (df). This implies replacing the log constant HR, β, in equation (5) by a flexible function of time $\beta(t)$. Quantin and collaborators [Quantin et al., 1999] provide an example of an application of this methodology in cancer epidemiology and illustrate its advantages over conventional PH regression. The 4-df likelihood ratio test, comparing the fit of the conventional 1-df proportional hazards model and the flexible 5-df regression spline model [Abrahamowicz et al., 1996] was used to verify the null hypothesis that the hazard ratio between those exposed and those not exposed did not change with increasing time since start of exposure (for manuscript II) or since start of infection (for manuscript III). Tests for linearity were also assessed with this method. All hypotheses were tested at the 0.05 significance level, using RRR, a customized software package for flexible models, provided by Dr. Abrahamowicz.

4.4.3.4 Cox regression with Time-dependent Covariates

The Cox model not only enables us to consider factors that were measured at the beginning of follow-up but also allows us to update the risk status over time through the use of time-dependent variables [Cox, 1972; Holford, 2002]. Cox regression with time-dependent covariates assigns different values of the covariate to the same subject at different points in time. At any time t, $X_i(t)$ represents the updated, usually most recent covariate value. An important assumption of the extended (time-dependent) Cox model is that the hazard at time t depends on the value of $x_j(t)$, i.e. the covariate value up-dated, at that same time t [Kleinbaum, 1996]. This hazard rate may depend on both fixed-in-time covariates $X_i(t)$ and time-dependent covariates $X_j(t)$ and is given by equation (7).

$$(7)\ h(t,\mathbf{X}(t)) = h_o(t) \exp\ [\ \sum_{i=1}^{p_1}\beta_i X_i + \sum_{j=1}^{p_2}\delta_j X_j(t)\ ]$$

The δ_j's are the log HR for the time-dependent covariates and the interpretation is similar to parameters for fixed covariates [Kleinbaum, 1996]. Cox regression produces a single estimate for δ_j for each time-dependent covariate. However, because the exposure is time-dependent, its impact on the hazard rate, for individual subjects, is also time-dependent.

In this study, many of the putative risk factors evaluated in manuscripts II and III were time-dependent internal variables. The distinguishing aspect of internal variables is that internal characteristics or behaviours specific to an individual prompt their changes. The time-dependent repeated measures were incorporated into programming statements that captured the appropriate covariate values of subjects in each risk set (SAS® version 8). Every time a "case" had an event of interest (i.e. tested HPV positive or cleared an HPV infection), the updated covariate pattern of each subject contained in the risk set was used to estimate the model parameters. In particular, the last measure of risk behaviour prior to, or at the time of the event, was used.

4.4.4 Logistic regression

Multiple logistic regression, a statistical method used in manuscript IV, is appropriate for analyzing data with dichotomous outcomes, such as presence of a cervical lesion (yes/no)

or clearance of a cervical lesion (yes/no), and can simultaneously adjust for potential confounding variables. In this model, if Y is the probability of disease, then Y/(1-Y) represents the "odds" of developing the outcome, and the log odds of disease, or the logit, can be written as ln[Y/(1-Y)] [Holford, 2002]]. The log odds of disease as the dependent variable can then be expressed as a simple linear function of the independent predictor variables, as expressed in equation (8).

(8) $\ln[Y/(1-Y)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$

The coefficients obtained through logistic regression can be directly converted to an odds ratio ($\exp(\beta_i)$) that provides an estimate of the relative risk that is adjusted for confounding, as in equation (9).

(9) $OR(x_i) = \exp^{\beta i}$

The 95% confidence limits around this estimate of relative risk can be obtained using the regression coefficient and its standard error (equation 10).

(10) $\exp^{(\beta i + 1.96 SE\beta i)}$

4.4.5 Evaluation of risk behaviour change and HPV acquisition or clearance

When the models in manuscripts II and III were developed, the association between risk behaviour and HPV infection were first examined by simple Cox regression analysis. However, because there was no single independent exposure of primary interest in this project, the 10% change in parameter estimate criteria [Rothman & Greenland, 1998] was not considered as a practical option for selecting variables for the final multivariable model. Instead, all variables that had a p-value of less than 0.20 were included in the multivariable analysis, in addition to those that were presumed to be important covariates based on the literature. The Wald test was used to evaluate the significance of individual parameter estimates. Under the null hypothesis, the Wald test-statistic has an approximate

normal distribution and is calculated by dividing the parameter estimate (β) by its standard error (SE), as given in equation (11).

$$(11)\ Z = \frac{\beta_i}{SE(\beta_i)}$$

4.4.6 Missing values

4.4.6.1 Missing data from questionnaires

If data related to a specific individual variable for a given visit was missing, the "last observation carried forward approach" was adopted and available information from the closest prior questionnaire was used. If still unavailable then data from the visit immediately subsequent to the visit with missing data was used. In the occasional situation where there was no available previous or subsequent data to replace missing data, the most frequent response (mode) or the mean value for that specific missing variable at that specific visit was assigned, based on the distribution of the entire study population. The mode value was assigned to categorical variables and the mean value, or median (if the distribution was skewed), was assigned to quantitative variables. The frequency of missing values for each variable at enrollment, are presented in the left column of table 1.i (Apprendix I.V) with the variable description. The right column of table 1.i represents the recoded variables, including the recoded missing values. In general, missing values represented less than 4% of all responses for a given variable. Responses for variables on contraceptive use or history of STDs tended to have a greater proportion of missing values (~10%) because participant's had a tendency to respond to the one or two practices or STDs applicable to them and skip the remaining variables. These missing values were coded as "never" since last visit.

4.4.6.2 Missing HPV results

Missing HPV results generally occurred when a sample had inadequate DNA for analysis and was classified as β-globin negative. In the case of samples that were β-globin negative (or missing altogether), the next visit with an informative HPV result was used. When analyzing the data with survival analysis, a conservative approach was taken that interpreted the β-globin negative visit as "unchanged status". Thus, for the analyses of

acquisition, among those women who were HPV negative, a β-globin negative visit would result in the classification of the subject as "still at risk for a new HPV infection". Similarly, for the analyses of clearance, among HPV positive women, a β-globin negative visit would result in the classification of the subject as "still at risk for HPV clearance".

For the analyses of HPV status as a risk factor for incident LSIL, if a subject was never HPV positive (i.e. either HPV negative and/or β-globin negative) at any time leading up to and at the time of an LSIL, the participant was classified as HPV negative. When estimating the average duration of LSIL's according to HPV status, if a subject was never HPV positive (i.e. either HPV negative and/or β-globin negative) at any time leading up to and at the time of an LSIL then the subject was excluded from the analysis.

4.4.6.3 Missing cytology data from laboratory

For the analyses of HPV status and incident LSIL, if a subject had missing data about cytology status at a specific visit then the subject was classified as not having a cervical abnormality at that visit.

4.4.7 Bias and Confounding

Age was considered as an *a priori* confounder in manuscripts II, III and IV, because it is associated with HPV acquisition [Bauer et al., 1993; Ho et al., 1998], HPV persistence [Hildesheim et al., 1994; Ahdieh et al., 2001] and cervical disease. In addition, age was likely to be associated with many other aspects of risky behaviour that were being investigated and was therefore, included in all the statistical models. Race was adjusted for in the fourth manuscript, because the literature has suggested that cervical disease may be associated with certain genetic polymorphisms and variant HPV types that may be distributed differentially according to ethnicity [Beckman et al., 1994; Sjalander et al., 1995]. Nonetheless, it should be noted that adjusting for race is a very imperfect method of controlling for genetic make-up. Race was composed of four categories (White, Black, Asian, Hispanic) that are highly prone to misclassification since each category was a compilation of many different ethnic groups that shared similar racial profiles (but not necessarily similar genetic backgrounds). Consequently the association between HPV variants and LSIL may still be biased due to residual confounding of genetic make-up.

Apparent behaviour change could have resulted from genuine changes in risk or from inaccurate reporting. To try and detect potential misreporting of risk behaviour, frequency distributions of each variable were assessed and any value that was unusually large or small was verified and compared to previous and subsequent responses. This would have an impact only if misreporting resulted in outlier values when compared to the overall distribution. However, because the final models in manuscripts II and III included the categorized conversions of continuous variables, the impact of extreme outliers on the estimated regression model is reduced.

Selection bias was investigated by comparing the characteristics of subjects lost to follow-up and those retained in the cohort. Despite losing approximately 30% of the cohort to follow-up, when we compared the distribution of exposures of interest among those women who did and did not complete the study (table 4.3), there were very few relevant differences. Table 4.3 also describes the distribution of certain outcomes, at enrollment, for the women who did and did not complete the study.

Table 4.3
Distribution of selected variables at baseline among those women lost to follow-up and among those women who completed follow-up

Baseline characteristics	Original cohort n=197 n (%)	Completed study n=424 n (%)	P-value (LR- $\chi 2$)
Age			
17-20	55 (28.0)	131 (30.9)	
21-23	86 (43.7)	142 (33.5)	
24-26	31 (15.7)	71 (16.9)	
27+	25 (12.7)	80 (18.8)	0.061
Race			
White	16.4 (83.2)	349 (82.2)	
Asian	14 (7.1)	48 (11.4)	
Black	13 (6.5)	16 (3.8)	
Hispanic	6 (3.0)	11 (2.6)	0.185
Smoking status			
Never	120 (60.9)	253 (59.6)	
Former	26 (13.2)	72 (17.1)	
Current	51 (25.9)	99 (23.3)	0.438
Cumulative cigarette consumption			
Non-smoker	121 (61.4)	254 (59.9)	
<1 pack/day for one yr	28 (14.2)	73 (17.2)	
1+ pack/day for > 1 yr	48 (24.3)	94 (22.2)	0.583
Alcohol consumption:			
0 drinks/week in lifetime	68 (34.5)	156 (36.8)	
1-3 drinks/week in lifetime	60 (30.5)	128 (30.3)	
>3 drinks/week in lifetime	69 (35.0)	137 (32.3)	0.20
Age at first sexual encounter			
19+ years	60 (30.5)	115 (27.1)	
16-18 yeas	92 (46.7)	220 (52.0)	
<16 years	45 (22.8)	89 (20.9)	0.481

Table 4.3 (continued)

Baseline characteristics	Original cohort n=197 n (%)	Completed study n=424 n (%)	P-value (LR- $\chi 2$)
Number of lifetime sex partners			
1 lifetime partners	52 (26.4)	95 (22.3)	
2-4 lifetime partners	51 (25.9)	145 (34.2)	
4-9 lifetime partners	61 (30.9)	100 (23.5)	
10+ lifetime partners	33 (16.8)	84 (20.0)	0.061
Number of new sex partners in last 6 months (baseline)			
0 new partners in 6 mo.	80 (40.6)	177 (41.8)	
1 new partner in 6 mo.	67 (34.0)	135 (31.8)	
2+ new partners in 6 mo.	50 (25.4)	109 (25.7)	0.889
Frequency of vaginal sex			
<1/week in lifetime	29 (14.7)	81 (19.1)	
1-2/week in lifetime	89 (45.2)	176 (41.5)	
>2/week in lifetime	79 (40.1)	164 (38.7)	0.370
Oral contraceptive use:			
Never in lifetime	53 (26.9)	91 (21.5)	
Sometimes in lifetime	17 (8.6)	47 (11.1)	
Always in lifetime	127 (64.5)	279 (65.9)	0.289
Condom use:			
Never in lifetime	25 (12.7)	29 (6.8)	
Sometimes in lifetime	61 (31.0)	147 (34.7)	
Always in lifetime	111 (56.3)	245 (57.8)	0.062
History of Chlamydia			
Never in lifetime	190 (7)	396 (93.4)	
Ever in lifetime	7 (3.6)	25 (5.9)	0.280
History of Warts			
Never in lifetime	174 (88.3)	360 (84.9)	
Ever in lifetime	23 (11.7)	61 (14.4)	0.405
HPV status at visit 1:			
Negative	125 (63.5)	289 (68.2.0)	
Positive	60 (30.5)	111 (26.2)	
β-globin negative	12 (6.1)	24 (5.7)	0.504
SIL status at visit 1:			
Normal	191 (97.0)	410 (96.7)	
ASCUS	1 (0.5)	1 (0.2)	
LSIL	5 (2.5)	12 (2.8)	0.779
HSIL	0 (0)	1 (0.2)	

4.4.8 Sample size and statistical power

One of the main objectives of this study was to estimate the rate of HPV persistence over two years. The sample size calculation was based on data from the NIH cohort [Hildesheim et al., 1994] that indicated that the overall HPV persistence, without regard to viral type, represented 45% of all infections. Based on the HPV positivity rate (35%) observed in the previous study at the McGill university student health clinic [Richardson et al., 2000] the estimated rate of persistence (HPV positivity at two consecutive visits) in the Montreal University population was 16% (45% of 35%). In order to measure the rate of persistence with adequate precision (at most 20% error, which translates into an absolute deviation of 3.2%) using 90% confidence intervals, we needed a cohort of size 356 subjects. Therefore, a cohort of 600 women, or more, should enable adequate levels of statistical precision for the estimation of rates of persistence, after accounting for an attrition rate of 10% per visit.

Another key question concerning statistical power was related to the identification of determinants of persistence. Using the projection for rates of persistence, a cohort of size 600 should enable sufficient power (80%) to detect, at the 0.05 significance level, a doubling in risk (OR=2) of persistence due to a risk factor present in 25% of the subjects' histories. The power for testing associations of higher magnitude should be even greater for OR >2.5.

The proposed sample size of 600 women should also ensure reasonable power for the analysis of cytologic endpoints, based on the proportion of women with cytological abnormalities in the previous study conducted with the same population [Richardson et al., 2000]. There were 59 (12.1%) abnormal Pap smears in the previous cross-sectional study. Of these, 39 were ASCUS, 17 were LSIL and 3 were HSIL. Assuming that the ratios of ASCUS:LSIL (2:1) and ASCUS:HSIL (13:1) seen in the previous study remain the same, the estimated number of incident cytological abnormalities in this study was projected to be close to 40%, or 210 women (40% of the 530 women who will be cytologically negative at enrollment). Of these, one-third or close to 70 women were

projected to have an incident SIL, based on the proportions observed in the previous study.

On the basis of these assumptions, for analyses encompassing both prevalent and incident SIL, the statistical power should be as high as 90% to detect a doubling in risk for factors present in 33% of the subjects (e.g. overall HPV infection). For rare exposures, e.g. persistent HPV infection, which is estimated at 16%, adequate power (80%) will be attained only for associations OR > 2.5. For the analyses focusing on incident SIL only, the statistical power will be adequate (80%) to detect RR>=3 for a factor present in as few as 16% of the subjects (i.e. persistent HPV infection).

CHAPTER 5: MANUSCRIPT I

The natural history of type-specific human papillomavirus infections in female university students

5.1 Preamble

The objective of manuscript I was to describe the natural history of cervical HPV infections. There are only 2 studies, from very different populations, that have documented the distribution of incidence rates for type-specific HPV infections [Franco et al., 1999; Giuliano et al., 2002a]. The average duration of type specific HPV infections has been estimated in approximately 6 studies [Ho et al., 1998] of which the vast majority included prevalent cases when calculating time-to-clearance [Moscicki et al., 1998; Franco et al., 1999; Ahdieh et al., 2001; Woodman et al., 2001; Giuliano et al., 2002a]. None of these studies accommodated for co-infections in their analyses, when estimating the duration of grouped HR-or LR-HPV infections.

In this study, the incidence rate and average duration of type-specific HPV infections was estimated for 27 HPV types, of which the 10 most frequently occurring types are presented in the manuscript. This study was the first in the literature to estimate the duration of mutually exclusive incident high-risk or low-risk type-specific HPV infections. This strategy was thought to yield a less biased estimate of duration by avoiding the mixing of a pre-existing HPV type and a newly acquired HPV type in the definition of incident HR-or LR-HPV persistence.

The natural history of type-specific human papillomavirus infections in female university students

(Running title: Natural dynamics of HPV infections)

Harriet Richardson[1,2], Gail Kelsall[1,3], Pierre Tellier[3], Hélène Voyer[5], Michal Abrahamowicz[2], Alex Ferenczy[1,4], François Coutlée[5] and Eduardo L. Franco[1,2]

Affiliation of authors: (1) Departments of Oncology, (2) Epidemiology & Biostatistics, (3) Family Medicine, and (4) Pathology, McGill University, Montreal, Canada; (5) Laboratoire de Virologie Moleculaire, Centre de Recherche et Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal, Montreal, Canada.

Correspondence to: Prof. EL. Franco, Department of Oncology, McGill University, 546 Pine Avenue West Montreal, QC, Canada H2W 1S6
(phone: 514-398-6032, e-mail: eduardo.franco@mcgill.ca).

Published in: Cancer Epidemiology, Biomarkers & Prevention. 2003;12(6):485-90

Abstract

Objectives: Little is known about the average duration of type-specific human papillomavirus (HPV) infections and their patterns of persistence. The objectives of this study were to evaluate the rate of acquisition and clearance of specific HPV types in young women. *Methods:* Female university students (n=621) in Montreal were followed for 24 months at 6-month intervals. At each visit a cervical specimen was collected. HPV DNA was detected using the MY09/MY11 PCR protocol followed by typing for 27 HPV genotypes by a line blot assay. The Kaplan-Meier technique was used to estimate the cumulative probability of acquiring or clearing an HPV infection considering types individually or in high or low risk groups defined by oncogenic potential.

Results: Incidence rates were 14.0 cases per 1000 women-months (95% CI: 11.4, 16.3) and 12.4 cases per 1000 women-months (95% CI: 10.4, 14.8) for acquiring HR-and LR-HPV infections, respectively. The 24-month cumulative rates of acquisition were highest for HPV 16 (12%), HPV 51 and HPV 84 (8%). Of the incident infections, HPV 16 was the most persistent (mean duration of 18.3 months), followed by HPV 31 and HPV 53 (14.6 and 14.8 months, respectively). HPVs 6 and 84 had the shortest mean duration time, at less than 10 months. The mean durations of incident, same-type LR- or HR-HPV infections were 13.4 months and 16.3 months, respectively.

Conclusion: While the majority of episodes with a type-specific HPV infection cleared within 2 years, there were many women who were either re-infected with a new HPV genotype or presumably experienced re-activation of their initial infection.

Introduction

While there is conclusive evidence that cervical human papillomavirus (HPV) infections are a necessary cause of cervical cancer [Bosch et al., 1995; Giannoudis & Herrington, 2001], the discrepancy between the high frequency of HPV infections in young, sexually active women and the relatively low occurrence of cervical lesions in the same population, suggests that HPV is not a sufficient cause for cervical neoplasia [Moscicki et al., 2001]. There is evidence that most HPV infections are transient and only women who harbour a persistent HPV infection are likely to develop a cervical lesion [Moscicki et al., 1993; Hildesheim et al., 1994]. However, there have been few studies designed to investigate the dynamics of HPV clearance or persistence. Describing the average duration of infection will be of great importance in establishing a clinically relevant definition of a persistent HPV infection that could be used for cervical screening and HPV vaccination studies [Bosch et al., 2002].

In 1996, we began a prospective cohort study of the natural history of HPV infection and cervical neoplasia in a population of young university students, in Montreal, Canada to study the rate of acquisition and clearance of specific HPV types in this population and to investigate risk factors for persistent HPV infections. This paper presents the descriptive epidemiologic results on the dynamics of acquisition, loss and persistence of type-specific HPV infections.

Materials and Methods

Subjects

Female students attending either the McGill or the Concordia University Health Clinic were invited to participate if they intended to be in Montreal for the next two years and had not required treatment for cervical disease in the last 12 months. Recruitment was initiated in November 1996 and accrual was completed in January 1999. All eligible women were asked to return to the clinic every 6 months over a period of 2 years, for a total of 5 visits. The study protocol was approved by the Research Ethics Boards of McGill University and Concordia University. At each visit, a questionnaire was completed and endo- and ectocervical cells from the uterine cervix were collected with

two Accelon cervical biosamplers (Medscand Inc., Hollywood, Fla.). A Pap smear was prepared from the first sampler.

HPV DNA detection

Preparation of the cell suspensions for HPV DNA testing has been described in detail elsewhere, with the use of QUIAamp columns (QUIAGEN Inc., CA. USA) for DNA purification [Coutlée et al., 1999]. Five µl of DNA was first amplified for β-globin DNA with PC04 and GH20 primers to demonstrate the absence of inhibitors and the integrity of processed DNA [Bauer et al., 1991; Coutlee et al., 2002]. β-globin-positive specimens were further tested with the L1 consensus HPV primers MY09/MY11 and HMB01 and the line blot assay (Roche Molecular systems, CA) for the detection of 27 genital HPV genotypes [Bauer et al., 1991; Gravitt et al., 1998]. HPV types were analyzed individually or in groups according to their oncogenic classification. High-risk HPV-(HR-HPV) types included those genotypes that are most frequently found in cervical tumours: HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68. All other individual types that were identified with the line blot assay were classified as low-risk HPV (LR-HPV) types [Bosch et al., 1995]: 6, 11, 26, 40, 42, 53, 54, 55, 57, 66, 82, 83 and 84 and MM9 [Gravitt et al., 1998].

Statistical Methods

The estimates of the incidence rate for a given genotype only included women at risk of acquiring that given genotype, so that women with a prevalent infection (at enrollment) for a specific HPV type were excluded from that specific risk set. Patterns of type-specific HPV positivity were described by comparing the overall number of visits positive for a specific HPV type, allowing for intermittent negative results, to the number of consecutive visits with the same HPV type.

The Kaplan-Meier technique [Kaplan & Meier, 1958] was used to estimate the cumulative probability of acquiring a specific HPV type or grouped-type infection (HR-HPV, LR-HPV) as a function of the length of follow-up for each HPV type or grouped infection, among women who were negative for the specific genotype or HPV group at

baseline. The Kaplan-Meier method was also used to estimate the proportion of women who remained positive for a specific (incident) HPV type or grouped-type-specific infection by considering their index infection, when first detected, as time zero. Type-specific prevalent infections were not included in the analysis for clearance. Thus, for a woman with a prevalent infection, her index infection was the longest enduring type-specific infection that was newly acquired after enrollment. Time to an event was defined as the time until the first visit when a subject was no longer HPV positive (for a given type). Subjects with both HR and LR types in the index visit were assigned to either the HR group or the LR group, based on the type with the longest duration. If there was a tie, the index visit was assigned to the HR-HPVgroup. In case of a tie within the high-risk group or low-risk group, all longest persisting infections had to have cleared to be given "clearance" status. Subjects were censored at their last visit. The median and mean duration of infection was estimated directly with the actuarial (Kaplan Meier) method. The Kaplan Meier technique was also used to evaluate clearance of an overall HR-HPVor an overall LR-HPV infection. In this second strategy (that simulated an analysis that would have been conducted with results from using a Hybrid Capture system), a HR- or LR-HPV infection was only considered to have cleared when a woman was no longer positive for any HR-HPVs or any LR-HPVs, respectively.

Results

A total of 635 women were initially recruited into the study. However, 13 subjects (2.0%) withdrew before completing their first questionnaire, along with one woman who had a β-globin negative sample at visit 1 and did not return for visit 2. There were a total of 2650 completed visits at the time of this analysis (mean of 4.3 visits/subject) and 2570 (97.6%) of the cervical specimens were suitable for HPV DNA testing. Women with a β-globin negative result were not excluded from the analyses, instead, the next visit with an informative HPV result was used. Loss to follow-up was approximately 10% per visit with approximately 90% of the participants returning for visit 3 (12 months) and 67.5% of the cohort returning for visit 5 (24 months), thus, contributing a total of 13,353 woman-months of follow-up (mean of 21.5 months follow-up/subject). The average time-interval between visits was normally distributed with the majority of women returning

within 5 to 7 months of their previous visit. The mean age was 23 years (median 21, range 17-42) and 45% of the women had ≥ 5 lifetime sexual partners. The majority of women (81%) described themselves as Caucasian, 60% of the participants had never smoked and 24% were current smokers.

Table 5.1 shows the prevalence at baseline and incidence for the most common HPV genotypes and grouped-type infections. The prevalence of HR-HPV infections was 21.8% and was 14.8% for LR-HPV infections. There were 327 women who had HPV detected at one or more visits during the study, and of those women, 124 (38%) had co-infections with a HR- and a LR-HPV type at the same visit. The 3 most common HPV types at enrollment were HPV-16 (7%), HPV-53 (4.3%) and HPV-84 (3.8%). HPV-16 (5.2 per 1000), HPV-84 (3.7 per 1000) and HPV-51 (3.4 per 1000) were the most frequent newly acquired infections, with all incidence rates expressed per month. The incidence rates for HR-HPV and LR-HPV infections were very similar (14.0 per 1000 and 12.4 per 1000, respectively). The cumulative rate for new HPV infections are shown in figure 5.1, and were 18.0% (95%CI: 14.1-21.9) at 1 year and 36.4% (95%CI: 31.3-41.5) at 2 years. The equivalent rates for HR-HPV were 12.7% (95%CI: 9.6, 15.8) and 29.0% (95%CI: 24.4, 33.4), and for LR-HPV they were 13.4% (95%CI: 10.4, 16.4) and 23.7% (95%CI: 19.7, 27.7) (figure 5.1).

Table 5.2 shows the pattern of persistence in the cohort. Of those women with 2 or more positive visits with the same HPV type, over 80% harboured these infections at consecutive visits, regardless of the number of positive visits and infection type group. This suggests that persistence tended to be above the levels of viral load that can be effectively detected by standardized cervical sampling coupled with a validated PCR assay.

The most persistent infections, with median retention times greater than 1 year, included those with types HPV-31 (20.0 months), 16 (19.4 months), HPV-54 (16.8 months) and HPV-53 (13.9 months) (table 5.3). The median retention times for the types that cleared most rapidly were HPV-6 and HPV-84 (~6.5 months). The median time for clearance of

an incident (type-specific) HR-HPV infection (13.2 months) was slightly higher than the median time to loss of an incident (type-specific) LR-HPV infection (12.3 months). Approximately 53 to 56% of the women with an incident LR- or HR-HPV infection, respectively, still remained positive after 1 year of follow-up. The mean duration of most incident infections were very similar to the median estimates, although HPV 16 was the most persistent type (mean=18.3 months), followed by HPV 53 (14.8 months) and HPV 31 (14.6). HPV types 6 and 84 had the shortest mean duration times. The mean time for clearance of an incident (type-specific) LR-HPV or HR-HPV infection was close to 13 or 16 months, respectively. The mean duration of an overall LR- or HR-HPV episode (not necessarily type-specific persistent) was between 16 and 17 months, respectively. After one year of follow-up, among women with an incident HPV infection, approximately 59 to 61% of the women remained positive for an epidsode with any LR- or HR-HPV types, respectively.

Discussion

We decided to present estimates of time to clearance for incident infections only, rather than pooling both prevalent and incident cases. Calculating the average duration times for both prevalent and incident infections could result in an overestimation of duration of an infection because prevalent cases could over-represent persistent infections at any point in time. We considered two definitions for HPV clearance when HPV types were classified into high-risk or low-risk groups. There were many instances when a participant had an infection with more than one HR-or LR-HPV type at the same visit or at a later visit. Therefore, the first definition was based on mutually exclusive high-risk or low-risk groups and only the most persistent HPV type was included in the analysis. If a woman had an equally persistent HR-and LR-HPV infection she was only included in the high-risk group, and a woman had to have cleared the longest persisting type before her HPV infection was considered cleared. This approach had the advantage of not mixing a pre-existing type and a newly acquired type in the definition of persistence (or in this case, clearance). However, this strategy also has its limitations since we selectively chose the most persistent HR or LR type-specific infection from each subject (with an incident HPV infection) and then estimated a global average duration for any HR-or any LR-HPV

infection. This method may lead to an over- or an under-estimation of the average duration of specific HPV group infections. For example, if our cohort happened to have an overrepresentation of one specific HR-HPV type that happened to persist, on average, for a longer duration than other high-risk types, then this analysis would tend to inflate the overall duration of HR-HPV infections. Therefore, this approach will lead to altered estimates of duration based on the type-distribution of HPV infections in a particular population. In our cohort, the median duration of the most (type-specific) persistent (incident) HR-HPV infection (13.2 months) was slightly higher than the median duration of a LR-HPV infection (12.3 months), but the confidence intervals of the two estimates overlapped considerably.

In our second definition of HR- or LR-HPV clearance, a woman had to have been negative for any high-risk, or any low-risk type, subsequent to any incident high-risk or incident low-risk infection at the previous visit, respectively. Thus, the concept of persistence could include a pre-existing (incident) high-risk (or low-risk) type infection mixed with a newly acquired (different) high-risk (or low-risk) type infection, provided there was not an intermittent visit that was high-risk (or low-risk) HPV negative. While this approach does not directly capture HR-or LR-HPV persistence with the same type, it does attempt to describe the dynamics of an infection that may contain more than one high-risk (or low-risk) type infection that is only resolved when all high-risk (or low-risk) types are cleared. It also simulates the strategy used by the majority of other researchers who have published data on the median duration of HPV infections [Ho et al., 1998; Moscicki et al., 1998; Franco et al., 1999; Woodman et al., 2001].

Given the large number of participants with more than one HR-or LR-HPV type infection throughout the study it was not surprising that the estimated median duration of an overall (incident) HR-HPV (16.6 months) or an overall (incident) LR-HPV (14.7 months) episode was longer than the estimated duration of the longest HR-or LR-HPV (type-specific) infection (13.2 months and 12.3 months, respectively). Most other studies have observed that the median duration of low-risk infections are less than 5 months while the median duration of HR-HPV infections are usually twice as long (8 to 10 months)

[Franco et al., 1999; Giuliano et al., 2002a]. While our second definition of HPV clearance highlights the high frequency of infection and re-infection with the same or different types (within the same oncogenic group), contrary to other reports [Franco et al., 1999; Giuliano et al., 2002a], our results suggest that the average duration of newly acquired high-risk or low-risk types do not differ substantially. However, most of the studies that have been able to estimate the duration of HPV infections have generally evaluated clearance of prevalent or mixed prevalent and incident infections [Moscicki et al., 1998; Franco et al., 1999; Woodman et al., 2001; Giuliano et al., 2002a], and only a few had assays for detecting a substantial number (>10) of low-risk types [Ho et al., 1998; Franco et al., 1999]. These different design issues could affect the estimates of duration of grouped HR-and LR-HPV infections.

Another explanation for our findings of a similar average duration for incident HR-HPV and LR-HPV infections is that there may not have been enough follow-up time to reveal the real average length of time to clearance. Only half of the women had a detected cleared incident HPV infection, versus 70% of those with LR-HPV (data not shown). The remaining women with an incident HPV infection were censored, i.e., possibly persistent. When we looked at clearance of prevalent HPV infections only, the majority of those prevalent HPV episodes cleared before the last follow-up visit (data not shown). Among the latter, HR-HPV infections persisted for an average of 19.5 months (95% CI: 16.9, 22.1) while LR-HPV infections were cleared, on average, within 16 months (95%CI: 13.7, 18.0).

While some recent studies have shown the median duration of new or prevalent HPV infections to be less than 10 months in young and middle aged women [Ho et al., 1998; Franco et al., 1999], distinguishing between an infection that has truly resolved and a false negative test result due to poor sampling, low levels of virus, or insensitive measurement tests is very difficult [Woodman et al., 2001]. As a result the clearance rate may be somewhat overestimated, while the frequency of persistent infections may be underestimated. At least one study [Woodman et al., 2001] observed that the median duration for an HPV infection (not necessarily type-specific) was greater than one year

(13.8 months) and when Moscicki and collaborators [Moscicki et al., 1998] considered various definitions of clearance, by modeling different number of consecutive HPV negative tests since the last HPV positive test, the median duration of the infection increased as the definition of clearance became more conservative. The authors concluded that it took approximately 15 months for 50% of the women in their study to clear a prevalent HPV infection, conditional on 3 consecutive negative HPV tests (but not necessarily a type-specific HPV infection). The median duration for any incident HPV episode in our cohort was 17.3 months (12.8, 21.7). While misclassification of HPV status may have occurred in our study, our results suggest that there were few false negative results, since very few women with a persistent type-specific infection had an intervening visit with a negative test result, and more than 80% of same-type persistent infections occurred during consecutive visits.

The time interval between visits can also influence the assessment of persistence. With a shorter interval between visits the clearance time for HPV episodes would appear earlier as would time of acquisition of a new HPV infection. The shorter interval may lead to improved precision of clearance time but would not necessarily change the estimate of mean duration. For practical purposes we opted for 6-month intervals between testing opportunities since the interval between HPV tests should be consistent with existing clinical guidelines for monitoring cervical cytological abnormalities, and is currently defined by most practice standards at 6-month intervals. In addition, between-test intervals of 6 months are more coherent with the biological rationale for using persistent HPV infection as an outcome in trials of HPV vaccine efficacy because it allows for the onset of induced immunity in clearing immediate post-vaccination transient infections. With short testing intervals such infections could mistakenly be interpreted as persistent and be counted as vaccine failure events, a scenario that would lead to a biased estimate of the vaccine efficacy.

Prevalence of cervical HPV infections has been investigated in numerous studies [IARC, 1995]. Nonetheless, the geographical variation in type distribution has not been extensively documented, except for HPV 16, which appears to be the most frequently

occurring type in most countries [Wheeler et al., 1993; Hildesheim et al., 1994; Ho et al., 1998; Liaw et al., 1999; Franco et al., 1999; Woodman et al., 2001]. Our results are in agreement with these previous studies, with a point prevalence of HPV 16, at enrollment of 7% in the whole population or 24% of all HPV positive samples at baseline. HPV 53 appears to be very prevalent in different populations including ours, but the high prevalence of HPV 84 (previously MM8) in our group differs from other recent studies presenting type-specific prevalence from cohorts with a broader age range, which show HPV 84 to be rare [Liaw et al., 1999; Franco et al., 1999; Lazcano-Ponce et al., 2001]. Among those women who were HPV negative at baseline in our cohort, approximately 36% were infected with HPV at some time during the 2-year period of the study. Recent longitudinal cohort studies have shown the 36-month cumulative incidence rates for acquiring any new HPV infection to range from 43% to 51% [Ho et al., 1998; Woodman et al., 2001; Moscicki et al., 2001], among women in their early twenties and younger. The high incidence rates of HPV 16, 51 and 84 that were seen in our group have also been observed in a cohort of University students in New York [Ho et al., 1998] and young women attending gynecology clinics for routine screening in Arizona [Giuliano et al., 2002a]. As an important caveat, we generally considered the natural history of LR- or HR-HPV infections independent of a co-infection with a different HPV group. Co-infections at the same visit made up 38% of the overall infections in our cohort, of which the majority were a low-risk type accompanied by one or more high-risk types (data not shown).

Another limitation of the study was that only 27 HPV types could be detected with our system. Although the most frequent and important (in terms of oncogenic potential) HPV types have been included in our line-blot PCR assay, our results are not directly comparable to those of studies that could detect more than 35 types [Ho et al., 1998; Franco et al., 1999]. It is conceivable that one or more LR-HPV types not included in our probe set could have been present in our population and thus would have constituted false-negative results, although the extent of the bias is probably small because of the rarity of such types.

In conclusion, the natural history of cervical HPV infections in this cohort of university students is anything but static. There is frequent acquisition of both HR-and LR-HPV types. The median duration of the longest persisting newly acquired LR- and HR-HPV infections is between 12 and 13 months, respectively, with great variation in the average duration of type-specific infections (range of 6.3 to 20 months among the 10 most common types). While the majority of the incident type-specific HPV infections cleared within 2 years, there were also many women who were either re-infected with a different HPV type or presumably experienced re-activation of their initial infection. Results from ongoing variant analyses of HPV 16 and HPV 18 in our cohort will help us determine how many of those type-specific re-infections actually represent the same infection. Finally, whether co-infections influence the natural history of type-specific infections still needs to be further explored, although preliminary investigations suggest that while risk of acquisition may be higher among women with co-infections, persistence is not affected [Thomas et al., 2000; Rousseau et al., 2001; Liaw et al., 2001].

Acknowledgments

This study was funded by a grant from the Canadian Institutes of Health Research (CIHR) (MT-13649). HR is recipient of a Pre-doctoral Scholarship, MA of a CIHR Scientist Award, and EF of a Distinguished Scientist Award, all from the CIHR. FC is recipient of a Chercheur Boursier National from the Fonds de la Recherche en Santé du Québec. The authors are indebted to Suzanne Dumais and the staff at the McGill and Concordia Student Health Services who helped with subject accrual and follow-up. The authors also thank Roche Molecular Systems who generously supplied us with all the reagents necessary for HPV genotyping.

Table 5.1

Prevalence and incidence of infection with the most frequently detected HPV types and for groups according to oncogenicity

HPV type	Baseline Prevalence (%)	Number of incident cases	Women months of follow-up	Incidence rate (per 1000 woman–months) (95% CI)
6	2.7	29	12709	2.3 (1.5, 3.3)
16	7.0	62	11928	5.2 (4.0, 6.7)
18	3.1	24	12735	1.9 (1.2, 2.8)
31	2.6	21	12854	1.6 (1.0, 2.5)
39	1.0	247	13476	1.8 (1.1, 2.5)
51	2.9	43	12588	3.4 (2.5, 4.6)
53	4.3	31	12468	2.5 (1.7, 3.5)
54	2.7	32	12783	2.5 (1.7, 3.5)
56	2.6	19	12842	1.5 (0.9, 2.3)
84	3.8	46	12475	3.7 (2.7, 4.9)
Any HPV	**29.0**	**155**	**8151**	**19.0 (16.1, 22.3)**
HR-HPV	**21.8**	**131**	**9344**	**14.0 (11.4, 16.3)**
LR-HPV	**14.8**	**128**	**10299**	**12.4 (10.4, 14.8)**

Table 5.2

Persistence versus intermittency of HPV infections according to number of visits with the same type-specific infection

Number of visits with the same type-specific infection[1]	Number of subjects with same type high-risk HPV infections		Number of subjects with same type low-risk HPV infections	
	Any combination of positive visits	Positivity in consecutive visits only (%)	Any combination of positive visits	Positivity in consecutive visits only (%)
2	73	68 (93.2)	43	36 (83.7)
3	33	27 (81.8)	21	18 (85.7)
4	24	23 (95.8)	7	6 (85.7)
5	16	16 (100.0)	5	5 (100)

[1] *Note: These are mutually exclusive groups and refer to the exact number of women with 2, 3, 4 or 5 visits with a type-specific HR or LR HPV infection.*

Table 5.3

Different measures of the time to loss of an incident infection with specific HPV types and for grouped-type specific infections

HPV type	Number of cases (n)	Median retention[1] time (95%CI) in months	Mean retention[1] time (95%CI) in months	Proportion[1] (%) remaining positive at 1 year (95%CI)
HPV-6	26	6.4 (4.9, 7.8)	8.7 (6.8, 10.6)	42 (19, 65)
HPV-16	62	19.4 (11.4, 27.5)	18.3 (12.9, 23.7)	62 (46, 78)
HPV-18	25	9.4 (4.8, 14.0)	11.6 (8.8, 14.4)	40 (15, 65)
HPV-31	21	20.0 (13.4, 26.6)	14.6 (11.0, 18.1)	62 (35, 89)
HPV-39	24	8.0 (5.8, 10.1)	11.0 (7.0, 14.9)	32 (3, 61.9)
HPV-51	45	9.0 (7.7, 10.4)	10.5 (8.4, 12.7)	35 (14, 56)
HPV-53	31	13.9 (11.1, 16.8)	14.8 (11.4, 18.3)	62 (41, 83)
HPV-54	34	16.8 (8.0, 25.7)	13.2 (10.2, 16.1)	58 (34, 82)
HPV-56	19	8.4 (3.2, 13.6)	10.6 (7.9, 13.2)	40 (13, 67)
HPV-84	47	6.6 (6.0, 7.2)	9.9 (7.0, 12.8)	23 (7, 41)
HR-HPV[2]	**124**	**13.2 (10.2, 16.2)**	**16.3 (13.7, 18.9)**	**56 (44, 68)**
LR-HPV[3]	**73**	**12.3 (11.4, 13.5)**	**13.4 (11.4, 15.4)**	**53 (41, 65)**
Any HPV episode[4]	**155**	**17.3 (12.8, 21.7)**	**17.0 (15.1, 18.8)**	**62 (52, 72)**
HR-HPV episode[5]	**131**	**16.6 (14.5, 18.7)**	**17.4 (14.7, 20.1)**	**61 (51, 71)**
LR-HPV episode[6]	**128**	**14.7 (10.9, 18.4)**	**15.8 (13.3, 18.3)**	**59 (49, 69)**

[1] *Estimates from actuarial analysis using the Kaplan-Meier technique*
[2] *HR-HPV infections were grouped according to the longest persisting (incident) HR type-specific infection*
[3] *LR-HPV infections were grouped according to the longest persisting (incident) LR type-specific infection.*
[4] *Episode refers to consecutive visits with any type*
[5] *Episode refers to consecutive visits with high-risk (not necessarily type-specific)*
[6] *Episode refers to consecutive visits with low-risk group infection (not necessarily type-specific)*

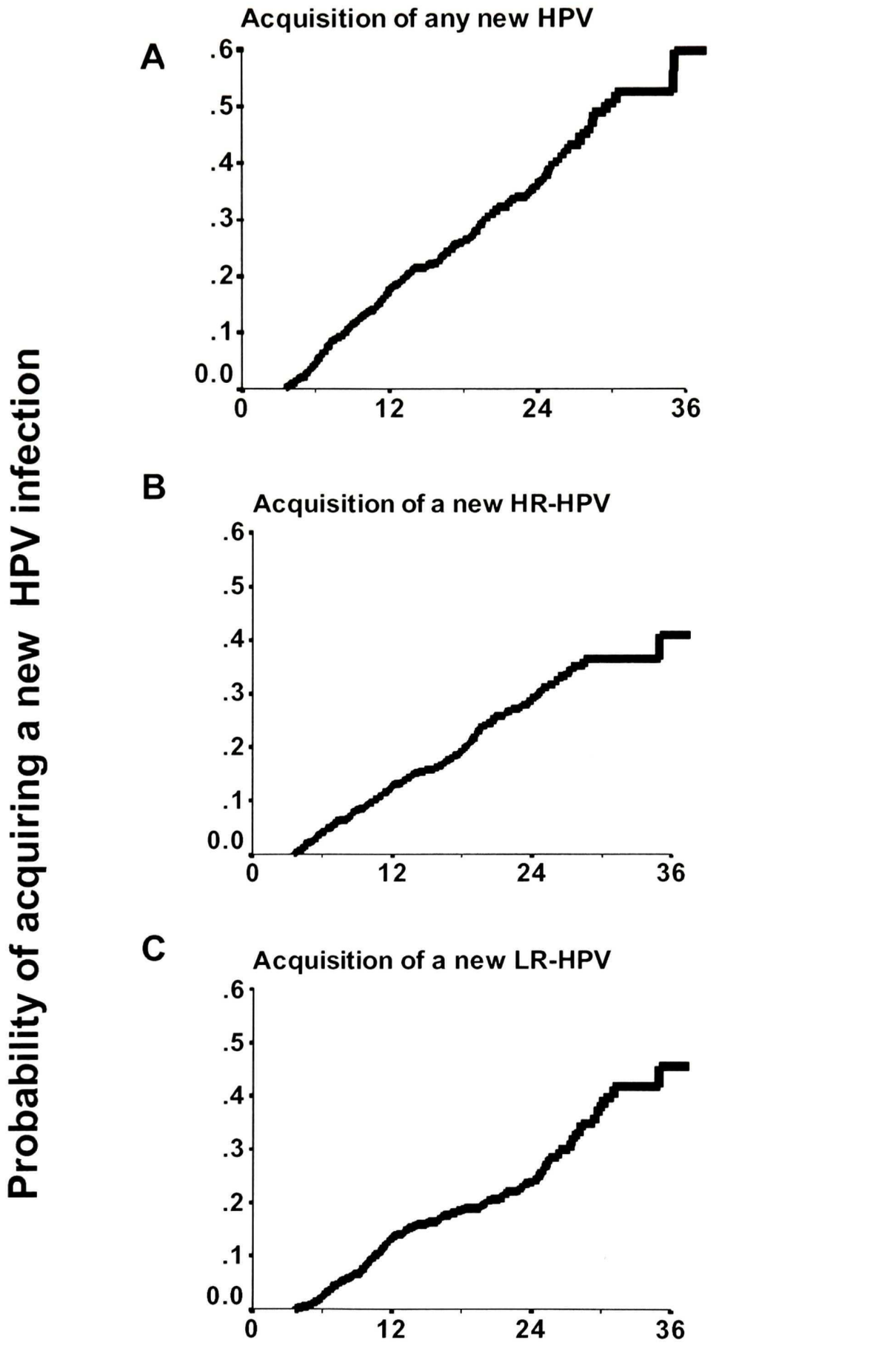
Acquisition of any new HPV
A
.6
.5
.4
.3
.2
.1
0.0
0
12
24
36
B
Acquisition of a new HR-HPV
.6
.5
.4
.3
.2
.1
0.0
0
12
24
36
C
Acquisition of a new LR-HPV
.6
.5
.4
.3
.2
.1
0.0
0
12
24
36
Probability of acquiring a new HPV infection
Time to detection of new HPV infection

Figure 5.1

Legend for Figure 5.1

Cumulative probability of incident HPV infections:

A) acquisition of any HPV type among women HPV negative at enrollment (n=420);

B) acquisition of a high-risk HPV infection among women HR-HPV negative at enrollment (n=460);

C) acquisition of a low-risk HPV infection among women LR-HPV negative at enrollment (n=498).

5.2 Additional analyses

Table II.i in Appendix II is an exhaustive list of the prevalence and incidence rates for all HPV types tested in the cohort. The three rarest HPV types in this cohort were HPV types 57, 31 and 40.

The 6, 12 and 24 month cumulative probabilities of acquiring one of the ten most common HPV types in this cohort are presented in table II.ii in Appendix II. In general, the annual risk for acquiring a new HPV type, among women who were negative for that type at baseline, appeared to be fairly constant over a two year period of follow-up.

The median and mean duration for all the ***incident*** type-specific HPV infections detected in the cohort are presented in table II.iii in Appendix II. Our data suggests that incident infections with HPV 55 clear the most quickly, while a newly acquired (low-risk) HPV 84 infection is as persistent, on average, as the high-risk HPV-16. However, because most of the type-specific infections were so rare, the estimates of average time-to-clearance for many individual HPV genotypes are very unreliable.

Average duration of the type-specific ***prevalent*** HPV genotypes, presented in table II.iv in Appendix II did not vary substantially from the estimated duration of the incident infections, presented in table II.iii.However, prevalent HR-HPV infections took nearly 20 months, on average to clear, while prevalent low-risk types took 16 months to clear on average. This difference was a little greater than the 1-month difference in duration between incident HR-and LR-HPV infections.

The time-to-clearance of incident and prevalent HPV infections, over a maximum of 36 months of follow-up time, are also represented graphically in figures II.i and II.ii, respectively, in Appendix II.

CHAPTER 6: MANUSCRIPT II

Co-factors for acquisition of low and high oncogenic risk cervical HPV infections in young women

6.1 Preamble

The factors that contribute to the occurrence of HPV infections have been studied extensively in cross-sectional studies. A few longitudinal cohort studies have recently been published confirming that age and sexual activity of a woman and her partner(s) are the strongest predictors of cervical HPV acquisition [Ho et al., 1998; Kruger-Kjaer et al., 2001; Moscicki et al., 2001; Winer et al., 2003]. Nonetheless, other co-factors are still thought to play a role in viral transmission, although there have not been any consistent findings.

Although a few studies have evaluated co-factors separately for prevalent high or low oncogenic risk HPV infections [Franco et al., 1995; Kjaer et al., 1997; Rousseau et al., 2000; Chan et al., 2002] no study to date has considered exclusively incident infections in the analyses. There may be some important biological distinctions between the two HPV groups and evaluating risk factors for any HPV infection may result in a diluted association, which may explain the inconsistent results in the literature. This manuscript fills a gap in the literature and provides results from two separate sets of analyses aimed at identifying determinants of incident low-risk (LR) and high-risk (HR) HPV infections. This manuscript also takes advantage of information collected repeatedly at each visit, so that change in risk behaviour over time can be properly addressed in the analyses.

In Manuscript I the natural dynamics of HPV acquisition and clearance for HPV types and oncogenic group was studied without considering contributing factors that may affect HPV incidence. In manuscript II the focus is on the determinants that can influence the rate of HPV acquisition.

Co-factors for acquisition of low and high-risk cervical HPV infections in young women

Harriet Richardson[1,2], Michal Abrahamowicz[2], Pierre Tellier[3], Gail Kelsall[1,3], Roxane du Berger[2], Alex Ferenczy[1,4], François Coutlée[5] and Eduardo L. Franco[1,2]

Affiliation of authors: (1) Departments of Oncology, (2) Epidemiology & Biostatistics, (3) Family Medicine, and (4) Pathology, McGill University, Montreal, Canada; (5) Laboratoire de Virologie Moleculaire, Centre de Recherche et Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal, Montreal, Canada.

Correspondence to: Prof. EL. Franco, Department of Oncology, McGill University, 546 Pine Avenue West Montreal, QC, Canada H2W 1S6 (e-mail: eduardo.franco@mcgill.ca).

Abstract

Background: Different risk factors for high and low-oncogenic risk HPV infections have been identified by epidemiologic studies. However, findings have been based mostly on cross-sectional study designs and not on cohort investigations based on repeated HPV DNA testing.

Objectives: To identify predictors of acquisition of high-oncogenic risk (HR) and low-oncogenic risk (LR) HPV infections in a prospective cohort study.

Methods: Female university students (n=621) in Montreal were followed-up for 24 months at 6-month intervals. At each visit a cervical cell specimen was collected. HPV DNA was detected using the MY09/MY11 polymerase chain reaction protocol and 27 HPV genotypes were identified by a line blot assay (Roche Molecular Systems Inc). Proportional hazards regression was used to estimate the crude and adjusted hazard ratios of acquiring a HR-or LR-HPV infection over time, for specific baseline and time-dependent covariates.

Results: A recent *Chlamydia* infection was predictive of a new high-risk (HR=9.6, 95% CI: 2.1, 43.9) or low-risk (HR=7.1, 95% CI: 1.5, 33.7) HPV infection. Increased duration of OC use was protective against acquisition of a LR-HPV infection (5+ years vs. non-users: HR =0.48, 95% CI: 0.23, 0.98). Washing after sexual intercourse was protective against acquiring a new HR-HPV infection (HR=0.54, 95% CI: 0.23, 1.3).

Conclusion: Co-factors were identified that appeared to influence the acquisition of a sexually transmitted HPV infection. Some distinct differences between the sets of predictors of low-risk and HR-HPV acquisition were apparent, but independent co-factors were also shared, including a recent *Chlamydia* infection.

INTRODUCTION

Cervical human papillomavirus (HPV) infection is currently one of the most common sexually transmitted agents and certain types of HPV are now considered to play a necessary role in cervical carcinogenesis [Bosch et al., 1995; Walboomers et al., 1999]. Numerous studies have investigated the role of demographic factors, sexual behaviour and other lifestyle choices in the detection of prevalent cervical HPV infection. The most important correlates of HPV infection that have been consistently identified by epidemiologic studies are age, number of lifetime partners [Rylander et al., 1994; Burk et al., 1996; Kruger-Kjaer et al., 2001], and age at first intercourse [Ley et al., 1991; Hildesheim et al., 1993; Bosch et al., 1994]. While HPV infection is primarily sexually transmitted, there is some evidence that infection with a prevalent high-oncogenic risk (HR) HPV type may be associated with a different set of risk factors than those with low-oncogenic risk (LR) HPV types [Franco et al., 1995; Kjaer et al., 1997; Rousseau et al., 2000; Richardson et al., 2000; Chan et al., 2002], thus associations with certain risk factors may be diluted if high and LR-HPV infections are analyzed together.

In addition, because prevalent infections detected in cross-sectional studies are a mixture of persistent and newly acquired infections, there is still a need to better understand which factors may facilitate the acquisition of a new HPV infection. This information will become increasingly important in the public health arena, since HPV testing, as an adjunct to Pap smear screening, is being considered as a viable option to improve the triage of cervical neoplasia [Manos et al., 1999; Ratnam et al., 2000; Solomon et al., 2001]. As women become more aware of the frequent occurrence of HPV infections, there will be more pressure to understand how to minimize risk of transmission in the future. To date, there are few prospective cohort studies that have assessed the determinants of incident HPV infection. Those that have, found that age at first intercourse was not an independent predictor of incident HPV infection [Ho et al., 1998; Kruger-Kjaer et al., 2001; Moscicki et al., 2001], and recent sexual behaviour was a stronger predictor of acquisition than number of lifetime partners [Ho et al., 1998; Elfgren et al., 2000; Kruger-Kjaer et al., 2001; Moscicki et al., 2001]. Results regarding putative associations between smoking, oral contraceptive use, condom use and alcohol

consumption and an incident HPV infection have been inconsistent [Ho et al., 1998; Kruger-Kjaer et al., 2001; Moscicki et al., 2001; Xi et al., 2002; Winer et al., 2003].

In 1996, we began a prospective cohort study of the natural history of HPV infection and cervical intraepithelial neoplasia in a population of university students, in Montreal, Canada. The objectives were to study the rate of acquisition and clearance of specific HPV types in this population, information that we presented in a previous report [Richardson et al., 2003], and to investigate risk factors for incident and persistent HPV infections. This paper reports the results of our analysis of determinants of acquisition of HR-and LR-HPV infections in this cohort.

METHODS

Subjects

Female students attending either the McGill or the Concordia University Health Clinic, in Montreal, Quebec, Canada, were invited to participate if they intended to remain in Montreal for the next two years and had not required treatment for cervical disease in the last 12 months. Recruitment was initiated in November 1996 and was completed in December 1998. The study protocol was approved by the Research Ethics Boards of McGill University and of the participating clinics.

Procedures

All eligible women were asked to return to the clinic every 6 months over a period of 2 years, for a total of 5 visits. At each visit, a questionnaire was completed and endo- and ectocervical cells from the uterine cervix were collected with two Accelon cervical biosamplers (Medscand Inc., Hollywood, Fla.). A Pap smear was prepared with the first sampler and the remaining cells along with cervical cells collected with a second sampler were used for HPV DNA testing. At enrollment, information from a detailed, self-administered questionnaire was obtained on potential risk factors such as socio-demographic variables, race, diet, smoking history, sexual behaviour, reproductive history, contraceptive and medical history and personal hygiene. An abridged (follow-up) questionnaire designed to measure changes in recent sexual practices and other lifestyle factors was completed at each subsequent visit.

HPV DNA detection

HPV DNA detection by polymerase chain reaction has been described previously [Richardson et al., 2003]. In brief, 5 μl of DNA purified with QUIAamp columns (QUIAGEN Inc., CA. USA) [Coutlée et al., 1999] was first amplified for β-globin DNA with PC04 and GH20 primers to check for inhibitors and to verify the integrity of processed DNA[Coutlee et al., 2002]. Specimens that were β-globin-positive were further amplified with the L1 consensus HPV primers MY09/MY11 and HMB01 with Amplitaq Gold (TaqGold; Perkin-Elmer-Cetus, Norwalk, CT) and amplicons were typed with the line blot assay (Roche Molecular systems, CA) as previously described for the detection of 27 genital HPV genotypes [Gravitt et al., 1998] [Coutlee et al., 2002]. HPV types were analyzed individually or in groups according to their oncogenic classification. High-risk HPV types included those genotypes that are commonly found in cervical carcinomas and their precursors: HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68. All other individual types that were identified with the line blot assay were classified as low-risk HPV types [Bosch et al., 1995]: 6, 11, 26, 40, 42, 53, 54, 55, 57, 66, 73, 82, 83 and 84 and MM9.

Statistical Methods

Cox's proportional hazards regression [Cox, 1972] was used to estimate the crude and adjusted hazard ratios of acquiring a new HPV infection over time, according to specific baseline characteristics and behaviours. Separate analyses were carried out for acquisition of incident HR- and incident LR-HPV infections. Variables with repeated measures were represented by their most recent value, using time-dependent covariates [Cox, 1972]. Women who were free of a HR-HPV infection (n=460) or free of a LR-HPV infection (n=498) at baseline and had at least one follow-up visit were included in the analysis for incident oncogenic or non-oncogenic HPV infections, respectively. Date of the first visit was used as the beginning of follow-up and time to event was defined as the time elapsed until the first visit when a relevant infection was detected. Observations on infection-free women were censored at the last follow-up visit.

The proportional hazards assumption was tested based on a flexible generalization of the Cox proportional hazards model [Abrahamowicz et al., 1996]. The flexible model allows the hazard ratio for selected exposures to change over time, according to an arbitrary function, the shape of which is estimated from the data using a quadratic regression

spline with 5 degrees of freedom (df). The 4-df likelihood ratio test, comparing the fit of the conventional 1-df proportional hazards model and the flexible 5-df regression spline model was used to verify the null hypothesis that the hazard ratio did not change with increasing time since start of exposure. All hypotheses were tested at the 0.05 significance level.

The main objective of the analyses was to evaluate predictors/co-factors of HPV acquisition other than sexual activity, such as smoking, alcohol consumption, oral contraceptive use, condom use and recent STD exposure. Variables, represented as a categorical variable, was first tested in a reduced model that adjusted for age and measures of sexual activity including age at first intercourse, lifetime sexual partners and number of recent sexual partners. Selected measures of tobacco and alcohol use, OC and condom use, recent STD exposure, and other variables for which at least one category yielded a p-value of less than 0.20 for the Wald test, were included in the final multivariable model and were mutually adjusted for each other. All analyses were stratified by clinic, since the effect of clinic on acquisition of HPV was not constant over time. The proportional hazards assumption was valid for all other variables of interest (data not shown).

RESULTS

There were 621 women initially enrolled in the study that completed a questionnaire and consented to HPV DNA testing. Of these, 578 (93%) returned for the second visit and were included in the analysis for acquisition of an oncogenic or non-oncogenic HPV infection, provided they did not have a HR- or LR-HPV infection, respectively, at enrollment. Women returned on average every 6.7 months, (95% CI: 6.4, 7.0: range 2-26 months). There were 460 women included in the analysis for HR-HPV acquisition, with 131 (28.5%) observed events. There were fewer prevalent non-oncogenic HPV infections, thus, 498 women were included in the analysis for LR-HPV acquisition, with 128 (25.7%) observed events.

The distribution of baseline values of selected risk factors for HPV remained fairly stable between the original cohort at baseline and the cohort members who completed all five

visits (table 6.1). The overall HPV point prevalence did increase slightly at each subsequent visit, and the proportion of HPV positive women at visit 5 (34.0%) was greater than the proportion of women who were HPV positive at enrollment (28.6). However, the overall distribution of low-grade or high-grade squamous intraepithelial lesions (LSIL and HSIL, respectively) did not change materially during 2 years of follow-up (table 6.1).

Table 6.2 presents results of Cox regression analyses of putative determinants of HR-and LR-HPV infection, with and without covariate adjustment for age and sexual activity. Age and markers of sexual activity (number of lifetime partners and new number of sexual partners since last visit), mutually adjusted for each other, were associated with acquisition of both LR- and HR-HPV infections. Age at first sexual debut was not significantly associated with an increased risk for a new HPV infection.

Determinants identified at baseline only are identified by "BL" and those variables measured repeatedly during follow-up are identified by "FUP" in tables 6.2 and 6.3. As shown in table 6.2, recent behaviour was a stronger predictor of HPV acquisition then past behaviour measured at baseline. Most estimates were attenuated when the baseline covariate values were used to assess an association with acquisition of HPV, instead of most recent values represented by time-dependent covariates. One exception was the direction of association between condom use and incident LR-HPV infections (table 6.2). Condom use appeared to be a strong risk factor for acquiring a LR-HPV infection when only the baseline values were used. However, the relationship shifted to a protective effect (though no longer statistically significant) when condom use was analyzed as a time-dependent covariate. The adjusted hazard ratios for most other risk factors and acquisition of either HR- or LR-HPV infections were not very different (table 6.2).

Table 6.3 shows the results of the final multivariable model that included selected dimensions of tobacco and and alcohol use, OC and condom use, recent history of a STD and other variables with p-values less than 0.20, in earlier models, adjusted for age and sexual activity. In general, while similar co-factors existed between the two HPV groups

defined by oncogenic potential, moderate differences in the set of predictors for HR-and LR-HPV infections were also observed. The risk of acquiring a new HPV infection decreased with increasing age, but only women over 27 years of age were significantly protected against acquiring a HR-HPV infection (p=0.04). Lifetime number of sex partners and recent number of new sexual partners remained strong predictors of HPV acquisition.

A recent diagnosis of a *Chlamydia trachomatis* infection (self-reported) was significantly associated with acquisition of any HPV infection (table 6.3). Women with a recent diagnosis of *Chlamydia* were nearly 10 times (HR=9.6, 95% CI: 2.1, 43.9) more likely to acquire a HR-HPV infection compared to women who did not have *Chlamydia.* The effect of a recent Chlamydia infection was slightly less for incident LR-HPV infections (HR= 7.1; 95% CI: 1.5, 33.7).

Tobacco use, when measured as number of packs smoked per day in lifetime (pack-years), was not associated with an HPV infection (table 6.3). However, women who consumed more than three alcoholic beverages per week were significantly more likely to acquire a HR-HPV infection than women who did not drink alcohol (HR=2.4, 95% CI: 1.2, 4.8). Weekly consumption of dairy products (at least one cup of milk or one serving of cheese) was significantly protective (HR=0.2, 95% CI: 0.1, 0.5) against the acquisition of a HR-HPV infection compared to women who did not consume dairy products. Similarly, women who washed within an hour after sex had substantially lower rates of acquisition as compared to those women who never washed after sex (table 6.3). In contrast, none of these variables had statistically significant effects on the acquisition of a LR-HPV infection (table 6.3). However, longer use of OCs significantly decreased a woman's risk of acquiring a new LR-HPV infection (5+ years vs. non-users: HR=0.48, 95% CI: 0.23, 0.98) and there appeared to be a dose-response relationship with length of OC use (table 6.3). Long-term exposure to OCs was not associated with acquisition HR-HPV infections.

DISCUSSION

The viral-host interaction is an intricate interplay at the molecular level and it appears that HR-HPV types are much more effective than LR-HPV types at initiating the cascade of events leading to cellular transformation and proliferation [Ho et al., 1995; Nobbenhuis et al., 1999]. Thus, it is conceivable that if there are biological differences that distinguish LR-HPV from HR-HPV infections, there may also be distinct sets of risk factors for acquisition and persistence associated with each type of HPV infection. Therefore, we decided to try and identify predictors for HR- and LR-HPV separately.

However, one of the limitations in this analysis was that there were very few women who only had a LR-HPV infection, throughout the study. Of the 621 women enrolled in the cohort, 259 (42%) tested positive for HR-HPV and 215 (35%) tested positive for LR-HPV at one or more visits [Richardson et al., 2003]. Yet, the majority of the women with LR-HPV infections (58%) had a co-infection with a HR-HPV type at the same visit. Because of the small number of exclusive LR-HPV infections, any woman with a LR-HPV infection was considered eligible for the subset analyses of predictors of non-oncogenic HPV acquisition. Therefore, the differences between the oncogenic and non-oncogenic HPV groups are slightly diluted since they are not mutually exclusive. Nonetheless, despite a possible dilution of the difference between putative risk factors for HR-versus LR-HPV infection, two different sets of determinants according to HPV oncogenicicity seemed to have emerged in this study.

There are a number of conceivable mechanisms that may either facilitate or hinder HPV acquisition, independent of sexual activity. Factors that may cause cervical irritation and affect the integrity of the cervical squamous epithelium such as the use of tampons or a current *Chlamydia* or HSV-2 infection could play a facilitating role in the transmission of HPV infection. Alternatively, other factors may hinder HPV transmission including precautionary practices such as the use of vaginal lubrication (to minimize vaginal abrasion), condoms and washing after sexual activity. Another possible mechanism for acquisition may be through the attenuation of the immune response to the virus, thus facilitating propagation, persistence and detection of a cervical HPV infection. Factors

that might have such endogenous or hormonal influences on the host immune system include tobacco metabolites, alcohol, oral contraceptives and certain dietary nutrients.

In general, recent sexual behaviour and other recent "risky" co-factors were more predictive of a new HPV infection then "risky" behaviours and characteristics measured at baseline. Capturing recent exposure profiles and changes in certain risk behaviours with time-dependent covariates may be a better method, then measures of exposure taken at baseline, to evaluate both the "facilitating" effect of a putatuive co-factor and the relevant time period of exposure to the co-factor, on acquisition of a sexually transmitted HPV infection. Presumably, the practice of using condoms regularly in the past should not influence current risk of acquiring a new infection, other then to act as a proxy for current practice of condom use. Therefore, time-dependent PH regression methods can allow the researcher to use the most up-to-date, best classified exposure information in the analysis. Nonetheless, it is important to also evaluate past (baseline) exposure so as to identify potential risk factors, such as duration of OC use or duration and intensity of tobacco use that may have a cumulative biological effect on the integrity of the squamous epithelium or the immune response and that may, in turn, influence risk of acquisition of HPV.

As expected, age and sexual activity were associated with HPV acquisition, irrespective of type. However, age at first sexual intercourse was not associated with acquisition of either type of HPV infection. It has been suggested that age at first intercourse may be a proxy for first exposure to HPV [Schiffman & Brinton, 1995; Deacon et al., 2000] and would, therefore, be a better predictor of a latent, persistent HPV infection rather than of a new infection that is more likely to be transient. Findings from our study and others [Ho et al., 1998; Kruger-Kjaer et al., 2001; Moscicki et al., 2001] suggest this to be the case.

A recent (since last visit) diagnosis of a *Chlamydia trachomatis* infection was significantly associated with acquisition of either LR- or HR-HPV infections. However, there were very few women in the study who had *Chlamydia*, so the estimates of effect lacked precision. Nonetheless, what was most interesting about these results was how

much more predictive of an incident HPV infection was a recent diagnosis, evaluated as a time-dependent covariate, compared to the baseline covariate that only measured lifetime exposure to *Chlamydia*. Past or current *Chlamydia trachomatis* infections have been associated with an increased likelihood of HR-HPV DNA detection in cross-sectional studies [Kjaer et al., 1997; Giuliano et al., 2002b] and cervical cancer, after adjusting for HPV status [Miller et al., 1991; Anttila et al., 2001]. In addition, a recent study observed that cervical inflammation was associated with an increased risk for high-grade neoplasia among women infected with HPV [Castle et al., 2001]. These results suggest that *Chlamydia* infections may act to irritate the cervical epithelium, thereby facilitating HPV transmission.

Current oral contraceptive use has been associated with prevalent or incident HPV infections in some studies [Ley et al., 1991; Hildesheim et al., 1993; Bauer et al., 1993; Giuliano et al., 1999; Moscicki et al., 2001] but not others [Ho et al., 1998; Elfgren et al., 2000; Ludicke et al., 2001; Kruger-Kjaer et al., 2001]. Results published from two other studies observed increased duration of OC use (>6 years) to be a strong risk factor for cumulative (12 month) HR-HPV infections among a cohort of middle aged women in Brazil [Rousseau et al., 2000], or observed that use of OC's 5-8 months, since prior visit, to be associated with the subsequent detection of non-prototypic HPV 16 variants, among university students in the USA [Xi et al., 2002]. In contrast to these results, we observed that women who used oral contraceptives for a total of 5 or more years in their life were significantly protected against acquiring a new LR-HPV infection. It is not clear why long-term OC use was associated with LR- but not HR-HPV acquisition. HPV DNA has been shown to contain hormone recognition elements and transformation of cells in vitro with viral DNA is enhanced by hormones [Pater et al., 1990; Auborn et al., 1991]. However, it is possible that the interaction between HPV types and cellular DNA is modified by differential hormonal responsiveness further downstream in the natural history of HPV infections, and OC use may only play a role in establishing a persistent infection or neoplastic progression. Nonetheless, measuring the association between OC use and HPV can be very challenging because OC use is so strongly correlated with sexual activity, and despite controlling for markers of sexual activity in the analysis

(including use of condoms), there may still be some residual confounding that may explain the protective effect of OC use against LR-HPV infections. Long-term OC use may be a proxy for a long-term monogamous relationship, or be measuring exposure to a "low-risk" sex partner.

It has been hypothesized that heavy drinkers (i.e. women who consume >14 alcoholic beverages/week) may possess higher circulating levels of estrogen and the vaginal epithelium may be more estrogen responsive, including the cervix [Reichman et al., 1993; Hankinson et al., 1995]. There were not many heavy drinkers in this cohort (e.g. >14 drinks/week) but we still observed a significant increased risk for acquisition of a HR-HPV infection among women who consumed more than 3 alcoholic beverages per week, compared to women who did not drink alcohol. Increased alcohol consumption was associated with acquisition and persistence of HPV in one cohort study of university students in New Jersey [Ho et al., 1998], but not another cohort of young women at high-risk recruited from family-planning and STD clinics in San Francisco [Moscicki et al., 2001]. Evidently, this potential risk factor needs to be further studied in light of potential confounding mechanisms before any firm conclusions can be made about alcohol consumption and risk of HPV infection.

While there is new evidence that a diet rich in vegetables is protective against a persistent HPV infection or preinvasive cervical lesions [Kjellberg et al., 2000; Sedjo et al., 2002a; Sedjo et al., 2002b], nutrition may also play a role in the steps leading to acquisition of an HPV infection. The majority of women (96.5%) in our study who consumed one serving of milk or cheese at least once a week were significantly less likely to acquire a HR-HPV infection compared to the remaining few who never consumed dairy products. It has been shown that HPV can integrate into the host DNA of several cervical cancer cell lines at fragile sites made susceptible to breakage by inadequate levels of folate [Popescu et al., 1987; Gallego et al., 1994]. It is possible that low levels of calcium, or other important micronutrients from dairy, may also facilitate the incorporation of HPV into the cervical epithelium in a similar manner. However, the protective effect of dairy products needs to

be corroborated in other studies before further speculation is made about the protective relationship of dairy and HPV transmission.

Hygienic practices, (including douching, use of tampons and washing after sex), have frequently been proposed as explanatory variables for observed differences in HPV prevalence, but reported associations have been inconsistent [Ho et al., 1998; Richardson et al., 2000; Rousseau et al., 2000]. The discrepancies may be due, in part, to little variation of certain hygienic practices between the participants in some of the studies. Nonetheless, in this study, women who washed regularly within an hour after sex were twice less likely to acquire a HR-HPV infection than women who never washed after sex. Similar findings were observed in this same population in an earlier study of prevalent HPV [Richardson et al., 2000].

Our study was not able to help identify a protective effect of condom use against HPV transmission. Despite the suggestive protective effect of recent condom use against acquisition of either LR- or HR-HPV infections, the point estimates were not statistically significant and there did not appear to be a dose-response with increasing regularity of condom use. Measuring the effect of condom use is extremely challenging, since condoms can serve two functions as a contraceptive barrier and/or an STD barrier. Because we did not make the distinction, in our questionnaire, between these two functions, we could not distinguish between women who used condoms throughout each entire sexual encounter, and those women who delayed the use of condoms towards the end of a sexual encounter, to prevent becoming pregnant. Furthermore, because we only measured frequency of condom use according to three levels (never, occasionally and always), a participant might have responded that she "always" used condoms with her partner even if they were only used 90% of the time. Consequently, exposure assessment of condom use may have been misclassified. Future studies that attempt to assess the potential protective effect of condoms need to develop a measurement instrument that can overcome these aforementioned limitations.

In summary, there did appear to be co-factors influencing the transmission of HPV, independent of sexual activity (as measured by number of lifetime sex partners and new number of recent sex partners). Moreover, despite some shared co-factors, there were some distinct differences between the predictors of low-risk and HR-HPV acquisition. While some of these co-factors are not modifiable, such as a recent *Chlamydia* infection, or at least not easily modifiable such as the use of oral contraceptives, other behaviours could be modified such as washing after sexual intercourse.

Acknowledgments

This study was funded by a grant from the Canadian Institutes of Health Research (CIHR) (MT-13649). HR is recipient of a Doctoral Scholarship, MA of a CIHR Scientist Award, and EF of a Distinguished Scientist Award, all from the CIHR. FC is recipient of a Chercheur Boursier National from the Fonds de la Recherche en Santé du Québec. The authors are indebted to Suzanne Dumais and the staff at the McGill and Concordia Student Health Services who helped with subject accrual and follow-up. The authors also thank Roche Molecular Systems who generously supplied us with all the reagents necessary for HPV genotyping.

Table 6.1

Distribution of selected variables at baseline among those women lost to follow-up and among those women who completed follow-up

Selected risk factors	Distribution at baseline among those women who completed follow-up N=424 n (%)	Distribution at baseline among those women lost to follow-up N=197 n (%)	LR~χ^2
Age			
17-20 years	131 (30.9)	55 (28.0)	
21-23 years	142 (33.5)	86 (43.7)	
24-26 years	71 (16.9)	31 (15.7)	
27+ years	80 (18.8)	25 (12.7)	*0.061*
Race			
White	349 (82.2)	164 (83.2)	
Black	16 (3.8)	14 (7.1)	
Hispanic	11 (2.6)	13 (6.5)	
Asian	48 (11.4)	6 (3.0)	*0.185*
Smoking status			
Never	253 (59.6)	120 (60.9)	
Former	72 (17.1)	26 (13.2)	
Current	99 (23.3)	51 (25.9)	*0.438*
Age at first intercourse			
19+ years	115 (27.1)	60 ((30.5)	
16-18 yeas	220 (52.0)	92 (46.7)	
<16 years	89 (20.9)	45 (22.8)	*0.200*
Number of lifetime sex partners			
1	95 (22.3)	52 (26.4)	
2-4	145 (34.2)	51 (25.9)	
5-9	100 (23.5)	61 (30.9)	
10+	84 (20.0)	33 (16.8)	0.061
HPV status (visit 1):			
Negative	289 (68.2)	125 (63.5)	
Positive	111 (26.2)	60 (30.5)	
β-globin negative	24 (6.1)	12 (6.1)	*0.504*
Cytological status (visit1)[1,2]**:**			
Normal	414 (97.6)	191 (97.0)	
ASCUS	1 (0.2)	1 (0.5)	
Low-grade SIL	8 (2.0)	5 (2.5)	
High-grade SIL	1 (0.2)	0 (0)	*0.779*

[1] *ASCUS=Atypical squamous cells of undetermined significance*
[2] *SIL=Squamous intraepithelial lesions*

Table 6.2
Determinants of acquisition of high-risk and low-risk HPV infections: crude & adjusted for age and sexual activity

	High oncogenic risk		Low oncogenic risk	
	Hazard Ratio (95%CI[a]) Crude	Hazard Ratio (95%CI) Adjusted[b]	Hazard Ratio (95%CI) Crude	Hazard Ratio (95%CI) Adjusted
A priori confounders				
Age				
17-20 years	REF	REF	REF	REF
21-23 years	1.04 (0.69, 1.57)	0.91 (0.57, 1.45)	0.96 (0.64, 1.45)	0.79 (0.51, 1.23)
24-26 years	0.95 (0.57, 1.60)	0.65 (0.35, 1.21)	0.72 (0.42, 1.24)	0.47 (0.25, 0.89)
27+ years	0.66 (0.38, 1.14)	0.39 (0.20, 0.77)	0.60 (0.34, 1.06)	0.39 (0.20, 0.76)
Age at first intercourse (BL)[c]				
19+ years	REF	REF	REF	REF
16-18 years	1.49 (0.96, 2.31)	0.82 (0.50, 1.32)	1.20 (0.78, 1.85)	0.72 (0.45, 1.15)
<16 years	1.76 (1.06, 2.92)	0.72 (0.39, 1.31)	1.51 (0.92, 2.48)	0.74 (0.41, 1.34)
No. lifetime sex partners (BL)				
1 sex partner	REF	REF	REF	REF
2-4 sex partners	2.06 (1.20, 3.55)	2.05 (1.17, 3.59)	2.09 (1.21, 3.61)	1.85 (1.05, 3.28)
5-9 sex partners	2.40 (1.38, 4.19)	3.02 (1.59, 5.77)	2.71 (1.55, 4.75)	3.30 (1.73, 6.31)
10+ sex partners	3.22 (1.84, 5.65)	4.14 (2.02, 8.47)	2.35 (1.29, 4.28)	2.51 (1.20, 5.24)
New no. sex partners (FUP)[d]				
0 new partners	REF	REF	REF	REF
1 new partner	2.42 (1.63, 3.60)	1.99 (1.33, 2.99)	2.32 (1.52, 3.54)	2.12 (1.38, 3.25)
2+ new partners	4.05 (2.050, 6.58)	2.90 (1.75, 4.80)	6.34 (4.03, 9.97)	5.12 (3.18, 8.23)

[a] 95% Confidence Interval; [b] Adjusted for age and sexual activity (number lifetime sexual partners and recent number of new sexual partners); [c] BL =Baseline variables; [d] FUP =Time-dependent variables

Table 6.2 (continued)
Determinants of acquisition of high-risk and low-risk HPV infections: crude & adjusted for age and sexual activity

	High oncogenic risk		Low oncogenic risk	
Putative co-factors	Hazard Ratio (95%CI[a]) Crude	Hazard Ratio (95%CI) Adjusted[b]	Hazard Ratio (95%CI) Crude	Hazard Ratio (95%CI) Adjusted
Race				
White	REF	REF	REF	REF
Black	0.75 (0.24, 2.35)	0.95 (0.29, 3.04)	1.09 (0.48, 2.49)	1.01 (0.43, 2.34)
Hispanic	0.86 (0.32, 2.32)	0.93 (0.34, 2.58)	0.37 (0.09, 1.50)	0.40 (0.10, 1.64)
Asian	0.86 (0.49, 1.49)	0.94 (0.52, 1.68)	0.71 (0.37, 1.36)	0.72 (0.37, 1.40)
Smoking status (BL[c])				
Never	REF	REF	REF	REF
Former	1.60 (1.03, 2.50)	1.12 (0.69, 1.80)	1.41 (0.89, 2.23)	1.27 (0.79, 2.06)
Current	1.62 (1.08, 2.42)	1.08 (0.70, 1.65)	1.54 (1.02, 2.32)	1.21 (0.78, 1.87)
No. cigarettes/day (BL)				
0-<1/day	REF	REF	REF	REF
1-5/day	1.13 (0.69, 1.86)	0.81 (0.48, 1.37)	1.36 (0.86, 2.15)	1.02 (0.62, 1.66)
>5/day	1.92 (1.30, 2.85)	1.38 (0.90, 2.14)	1.59 (1.04, 2.44)	1.58 (1.00, 2.50)
No. cigarettes/day (FUP[d])				
0-<1/day	REF	REF	REF	REF
1-5/day	1.04 (0.61, 1.77)	0.61 (0.35, 1.06)	1.98 (1.26, 3.10)	1.32 (0.83, 2.12)
>5/day	1.59 (0.99, 2.57)	1.04 (0.63, 1.71)	2.00 (1.21, 3.03)	1.76 (1.05, 2.95)
Pack-years smoked				
None	REF	REF	REF	REF
<1 pack-year	1.38 (0.87, 2.19)	0.83 (0.51, 1.35)	1.52 (0.97, 2.39)	1.12 (0.70, 1.79)
1+ pack-years	1.81 (1.22, 2.69)	1.38 (0.89, 2.13)	1.45 (0.95, 2.21)	1.34 (0.85, 2.13)
Alcohol consumption (BL)				
0 drinks/wk	REF	REF	REF	REF
1-3 drinks/wk	1.20 (0.78, 1.84)	1.24 (0.80, 1.92)	0.78 (0.49, 1.24)	0.81 (0.50, 1.32)
>3 drinks/wk	1.46 (0.97, 2.21)	1.05 (0.68, 1.64)	1.32 (0.89, 1.96)	1.04 (0.68, 1.59)

[a] 95% Confidence Interval; [b] Adjusted for age and sexual activity (number lifetime sexual partners and recent number of new sexual partners); [c] BL =Baseline variables; [d] FUP =Time-dependent variables

Table 6.2 (continued)
Determinants of acquisition of high-risk and low-risk HPV infections: crude & adjusted for age and sexual activity

	High-risk		Low-risk	
Putative co-factors	Hazard Ratio (95%CI[a]) Crude	Hazard Ratio (95%CI) Adjusted[b]	Hazard Ratio (95%CI) Crude	Hazard Ratio (95%CI) Adjusted
Alcohol consumption (FUP)				
0 drinks/wk	REF	REF	REF	REF
1-3 drinks/wk	2.66 (1.43, 4.94)	2.02 (1.07, 3.78)	1.07 (0.64, 1.77)	0.84 (0.50, 1.40)
>3 drinks/wk	3.69 (1.96, 6.96)	2.30 (1.19, 4.44)	2.22 (1.34, 3.65)	1.49 (0.88, 2.53)
Oral contraceptive use (BL[c])				
Never	REF	REF	REF	REF
Sometimes	1.20 (0.64, 2.25)	1.01 (0.52, 1.95)	0.73 (0.38, 1.42)	0.64 (0.32, 1.27)
Always	1.05 (0.68, 1.60)	0.91 (0.58, 1.42)	0.77 (0.51, 1.16)	0.65 (0.42, 1.00)
Oral contraceptive use (FUP[d])				
Never	REF	REF	REF	REF
Sometimes	1.61 (0.68, 3.79)	1.34 (0.56, 3.21)	2.40 (1.18, 4.89)	1.84 (0.89, 3.81)
Always	1.58 (1.09, 2.28)	1.56 (1.07, 2.28)	0.96 (0.67, 1.39)	0.94 (0.65, 1.37)
Duration of OC use				
0 years	REF	REF	REF	REF
<1 year	1.14 (0.67, 1.93)	0.97 (0.56, 1.67)	0.77 (0.46, 1.30)	0.63 (0.37, 1.06)
1-5 years	0.98 (0.62, 1.56)	0.88 (0.55, 1.43)	0.81 (0.52, 1.25)	0.68 (0.43, 1.08)
5+ years	1.20 (0.70, 2.08)	0.96 (0.50, 1.85)	0.65 (0.36, 1.17)	0.54 (0.27, 1.07)
Condom use (BL)				
Never	REF	REF	REF	REF
Sometimes	0.92 (0.45, 1.87)	0.63 (0.30, 1.32)	3.46 (1.07, 11.18)	3.39 (1.01, 11.34)
Always	1.32 (0.68, 2.53)	0.86 (0.43, 1.72)	3.93 (1.24, 12.42)	3.75 (1.15, 12.30)
Condom use (FUP)				
Never	REF	REF	REF	REF
Sometimes	1.08 (0.70, 1.63)	0.79 (0.50, 1.25)	1.09 (0.69, 1.71)	0.76 (0.48, 1.23)
Always	1.16 (0.78, 1.73)	0.74 (0.48, 1.15)	1.34 (0.89, 2.02)	0.87 (0.55, 1.38)
Foam use (BL)				
Never	REF	REF	REF	REF
Sometimes	0.71 (0.40, 1.27)	0.60 (0.33, 1.08)	0.79 (0.44, 1.40)	0.74 (0.41, 1.33)
Always	1.28 (0.62, 2.63)	1.45 (0.69, 3.02)	1.08 (0.50, 2.31)	1.35 (0.61, 2.96)

[a] 95% Confidence Interval; [b] Adjusted for age and sexual activity (number lifetime sexual partners and recent number of new sexual partners); [c] BL =Baseline variables; [d] FUP =Time-dependent variables

Table 6.2 (continued)
Determinants of acquisition of high-risk and low-risk HPV infections: crude & adjusted for age and sexual activity

	High-risk		Low-risk	
Putative co-factors	Hazard Ratio (95%CI[a]) Crude	Hazard Ratio (95%CI) Adjusted[b]	Hazard Ratio (95%CI) Crude	Hazard Ratio (95%CI) Adjusted
Foam use (FUP)				
Never	REF	REF	REF	REF
Sometimes	1.85 (0.93, 3.69)	1.12 (0.53, 2.39)	1.16 (0.47, 2.85)	1.31 (0.52, 3.33)
Always	0.60 (0.15, 2.43)	0.65 (0.16, 2.66)	0.27 (0.04, 1.96)	0.35 (0.05, 2.53)
Menstrual products (BL[c])				
Pads	REF	REF	REF	REF
Tampons	2.09 (1.21, 3.60)	1.34 (0.76, 2.39)	1.30 (0.73, 2.31)	0.89 (0.49, 1.62)
Pads & tampons	1.19 (0.72, 1.98)	1.04 (0.62, 1.75)	1.09 (0.67, 1.78)	0.82 (0.49, 1.36)
Wash after sex (BL)				
Never/rarely	REF	REF	REF	REF
Sometimes	0.87 (0.61, 1.25)	0.92 (0.64, 1.32)	1.08 (0.74, 1.57)	1.11 (0.75, 1.63)
Always	0.49 (0.21, 1.15)	0.39 (0.16, 0.92)	0.75 (0.35, 1.59)	0.83 (0.38, 1.79)
Wash after sex (FUP[d])				
Never/rarely	REF	REF	REF	REF
Sometimes	1.28 (0.90, 1.84)	1.06 (0.74, 1.52)	1.23 (0.85, 1.77)	1.00 (0.69, 1.47)
Always	0.53 (0.23, 1.23)	0.54 (0.23, 1.27)	0.76 (0.37, 1.54)	0.78 (0.38, 1.60)
Frequency of sex (BL)				
Sex <1/wk	REF	REF	REF	REF
Sex 1-2/wk	0.83 (0.46, 1.12)	0.82 (0.52, 1.29)	0.68 (0.44, 1.06)	0.81 (0.52, 1.27)
Sex 3+/wk	0.87 (0.53, 1.24)	0.87 (0.56, 1.34)	0.74 (0.48, 1.13)	0.81 (0.53, 1.24)
Frequency of sex (FUP)				
Sex <1/wk	REF	REF	REF	REF
Sex 1-2/wk	1.08 (0.68, 1.72)	1.19 (0.74, 1.90)	0.87 (0.56, 1.35)	0.85 (0.54, 1.32)
Sex 3+/wk	1.73 (1.15, 2.60)	1.65 (1.09, 2.50)	0.94 (0.62, 1.42)	0.92 (0.60, 1.40)

[a] 95% Confidence Interval; [b] Adjusted for age and sexual activity (number lifetime sexual partners and recent number of new sexual partners); [c] BL =Baseline variables; [d] FUP =Time-dependent variables

Table 6.2 (continued)
Determinants of acquisition of high-risk and low-risk HPV infections: crude & adjusted for age and sexual activity

	High-risk		Low-risk	
Putative co-factors	Hazard Ratio (95%CI[a]) Crude	Hazard Ratio (95%CI) Adjusted[b]	Hazard Ratio (95%CI) Crude	Hazard Ratio (95%CI) Adjusted
History of *Chlamydia* (BL[c])				
No	REF	REF	REF	REF
Yes	1.95 (0.99, 3.84)	1.47 (0.72, 3.01)	1.40 (0.62, 3.19)	0.88 (0.36, 2.12)
History of *Chlamydia* (FUP[d])				
No	REF	REF	REF	REF
Yes	6.75(1.62, 28.19)	7.35 (1.69, 32.01)	5.27 (1.28, 21.73)	4.48 (1.01, 19.86)
History of Herpes (BL)				
No	REF	REF	REF	REF
Yes	1.70 (0.63, 4.61)	1.27 (0.45, 3.57)	1.89 (0.70, 5.13)	2.06 (0.74, 5.71)
Vaginal irritation (BL)				
Never	REF	REF	REF	REF
1/yr	1.76 (1.04, 2.97)	1.88 (1.10, 3.23)	1.60 (0.93, 2.75)	1.51 (0.86, 2.65)
>1/yr	1.35 (0.75, 2.41)	1.46 (0.80, 2.67)	1.27 (0.71, 2.29)	1.07 (0.58, 1.97)
Vaginal irritation (FUP)				
Never	REF	REF	REF	REF
1-2/since last visit	0.88 (0.60, 1.28)	0.79 (0.54, 1.16)	1.22 (0.84, 1.78)	1.03 (0.70, 1.50)
>2/ since last visit	1.59 (0..96, 2.65)	1.41 (0.84, 2.36)	1.16 (0.65, 2.07)	0.94 (0.52, 1.68)
Dairy consumption				
Never	REF	REF	REF	REF
1+ servings/wk	0.34 (0.17, 0.67)	0.25 (0.12, 0.50)	0.56 (0.27, 1.15)	0.69 (0.32, 1.46)

[a] 95% Confidence Interval; [b] Adjusted for age and sexual activity (number lifetime sexual partners and recent number of new sexual partners); [c] BL =Baseline variables; [d] FUP =Time-dependent variables

Table 6.3
Final multivariable model for acquisition of high-risk and low-risk HPV infections

Variables	High -risk HPV acquisition		Low-risk HPV acquisition	
	Hazard Ratio	95% CI [a]	Hazard Ratio	95% CI
Age				
17-20 years	REF	REF	REF	REF
21-23 years	1.09	0.67, 1.78	0.84	0.53, 1.32
24-26 years	0.77	0.40, 1.49	0.53	0.28, 1.04
27+ years	0.44	0.21, 0.91	0.46	0.22, 0.96
Age of first sex encounter				
19+ years	REF	REF	REF	REF
16-18 years	0.86	0.52, 1.43	0.78	0.48, 1.27
<16 years	0.76	0.40, 1.43	0.77	0.42, 1.41
No. lifetime sex partners				
1 sex partner	REF	REF	REF	REF
2-4 sex partners	1.92	1.07, 3.45	1.75	0.98, 3.14
5-9 sex partners	2.40	1.21, 4.73	3.19	1.64, 6.20
10+ sex partners	3.12	1.41, 6.92	2.34	1.07, 5.15
New no. sex (FUP) [b]				
0 new partners	REF	REF	REF	REF
1 new partner	2.29	1.49, 3.51	2.20	1.39, 3.48
2+ new partners	3.42	2.00, 5.87	5.83	3.45, 9.85

[a] *95% CI= 95% Confidence Interval;* [b] *FUP =Time-dependent variables*

Table 6.3 (continued)
Final multivariable model for acquisition of high-risk and low-risk HPV infections

Variables	High -risk HPV acquisition		Low-risk HPV acquisition	
	Hazard Ratio	95% CI [a]	Hazard Ratio	95% CI
Pack-years smoked				
None	REF	REF	REF	REF
<1 pack-year	0.84	0.51, 1.38	1.01	0.67, 1.74
1+ pack-years	1.04	0.64, 1.68	1.07	0.65, 1.74
Alcohol consumption (FUP) [b]				
0 drinks/wk	REF	REF	REF	REF
1-3 drinks/wk	2.23	1.17, 4.25	0.83	0.49, 1.42
>3 drinks/wk	2.41	1.20, 4.83	1.58	0.92, 2.73
Duration of OC use				
0 years	REF	REF	REF	REF
<1year	0.94	0.53, 1.67	0.64	0.37, 1.10
1-5 years	0.80	0.48, 1.35	0.61	0.38, 0.99
5+ years	1.04	0.52, 2.07	0.48	0.23, 0.98
Condom use (FUP)				
Never	REF	REF	REF	REF
Sometimes	0.70	0.44, 1.11	0.69	0.42, 1.12
Always	0.70	0.44, 1.10	0.76	0.48, 1.22
Chlamydia (FUP)				
Never	REF	REF	REF	REF
Yes	9.57	2.09, 43.90	7.06	1.48, 33.68

[a] *95% CI= 95% Confidence Interval;* [b] *FUP =Time-dependent variables*

Table 6.3 (continued)
Final multivariable model for acquisition of high-risk and low-risk HPV infections

Variables	High -risk HPV acquisition		Low-risk HPV acquisition	
	Hazard Ratio	95% CI [a]	Hazard Ratio	95% CI
Dairy consumption				
Never	REF	REF	not included in final	
1+ servings/wk	0.23	0.11, 0.49	model	
Frequency of sex (FUP) [b]				
<1/wk	REF	REF	not included in final	
Sex 1-2/wk	1.32	0.81, 2.15	model	
Sex >2/wk	1.85	1.20, 2.87		
Wash after sex (FUP)				
Never/rarely	REF	REF	not included in final	
Sometimes	1.05	0.72, 1.52	model	
Always	0.52	0.22, 1.25		
Vaginal irritation (FUP)				
No	REF	REF	not included in final	
1-2 since last visit	0.71	0.48, 1.06	model	
>2/ since last visit	1.30	0.76, 2.20		

[a] *95% CI= 95% Confidence Interval;* [b] *FUP =Time-dependent variables*

6.2 Additional analyses

Tables III.i, III.ii and III.iii in Appendix III are the results from three PH multivariable regression models that were built to reflect a series of plausible biologic mechanistic models, adjusted for age, sexual activity.

The first model, "immune modifiers" (table III.i) included correlated variables that could facilitate HPV acquisition by impacting on the health of the cervical epithelial environment and the immune response, and included tobacco and alcohol use, dietary habits, and OC use. The parameter estimates for the variables of interest did not change materially from the adjusted estimates presented in table 6.2, once mutually adjusted in the "immune modifier" multivariable model (table III.i).

The second model, "cervical irritatnts" (table III.ii) included correlated variables that could act as cervical irritants and/or cause cervical inflammation that in turn could faciliatate HPV acquisition, such as a recent STD infection, high frequency of sex, use of tampons and general vaginal irritation. The parameter estimates for the variables of interest did not change materially from the adjusted estimates presented in table 6.2, once mutually adjusted in the "cervical irritants" multivariable model (table III.ii).

The third model, "protective factors" (table III.iii) included correlated variables that may have protective properties and hinder HPV acquisition, such as condom use, vaginal lubrication ("foam") and washing after sex. The parameter estimates for the variables of interest did not change materially from the adjusted estimates presented in table 6.2, once mutually adjusted in the "protective factors" multivariable model (table III.iii).

CHAPTER 7: MANUSCRIPT III

Modifiable risk factors associated with clearance of type-specific cervical HPV infections

7.1 Preamble

A very high proportion of women will have an HPV infection soon after initiating sexual relationships. The majority of women will eventually clear their HPV infection within 2 years. Nonetheless, a small fraction of these infections will persist. Approximately four studies have tried to elucidate determinants that contribute to a persistent HPV [Hildesheim et al., 1994; Brisson et al., 1996; Ho et al., 1998; Sedjo et al., 2002b]. Regardless of the particular study definition of HPV persistence, only one of these studies compared persistent infections to transient infections. However, the published results of this one study focused primarily on viral characteristics rather than modifiable risk factors [Ho et al., 1998].

The objective of manuscript III was to try and understand, by means of actuarial analysis techniques, what behaviours differed between those women with transient infections and those women who could not clear their type-specific HPV infection. Only one other study has used this approach to studying persistence [Moscicki et al., 1998], but the authors only considered time-to-clearance of <u>any</u> HR-or <u>any</u> LR-HPV episode, rather than clearance of a type-specific HR-or LR-HPV infection. Furthermore, this is the first study, to our knowledge, to accommodate changes in risk behaviour over time in the analyses.

In manuscript I the natural history of HPV clearance for type-specific HR-or LR-HPV infections was described without considering factors that may affect HPV clearance. In manuscript III the focus is on the determinants that can influence the rate of HPV clearance.

Modifiable risk factors associated with clearance of type-specific cervical HPV infections

Harriet Richardson[1,2], Michal Abrahamowicz[2], Pierre Tellier[3], Gail Kelsall[1,3], Roxane du Berger[2], Alex Ferenczy[1,4], François Coutlée[5] and Eduardo L. Franco[1,2]

Affiliation of authors: (1) Departments of Oncology, (2) Epidemiology & Biostatistics, (3) Family Medicine, and (4) Pathology, McGill University, Montreal, Canada; (5) Laboratoire de Virologie Moleculaire, Centre de Recherche et Departement de Microbiologie et Infectiologie, Hopital Notre-Dame du Centre Hospitalier de l'Université de Montréal, Montreal, Canada.

Correspondence to: Prof. EL. Franco, Department of Oncology, McGill University, 546 Pine Avenue West Montreal, QC, Canada H2W 1S6 (e-mail: eduardo.franco@mcgill.ca).

Abstract

Background: Previous findings regarding risk factors for HPV persistence, other than viral determinants, identified from prospective cohort studies have been inconsistent.

Objectives: Identify predictors of clearance of high- (HR) and low-risk (LR) HPV infections in a prospective cohort study.

Methods: Female university students (n=621) in Montreal were followed for 24 months at 6-month intervals. At each visit a cervical cell specimen was collected. HPV DNA was detected using the MY09/MY11 PCR protocol and 27 HPV genotypes were identified by the line blot assay (Roche Molecular Systems Inc). Proportional hazards regression was used to estimate the crude and adjusted hazard ratio of clearing a type-specific HR- (n=191) or LR-HPV (n=87) infection over time, according to specific baseline and time-dependent covariates.

Results: Daily consumption of vegetables appeared to protect women against a persistent HPV infection, independent of type. The use of tampons was significantly associated with an elevated risk of type-specific HR-HPV persistence, while regular condom use appeared to be protective against persistence of a low-risk HPV infection only.

Conclusion: Some proactive measures can be taken to increase the rate of HPV clearance, but the differences between the sets of predictors of LR- and HR-HPV clearance are not substantial.

INTRODUCTION

The identification of certain sexually acquired HPV infections as the necessary etiological agent for cervical cancer [Bosch et al., 1995; Walboomers et al., 1999] has helped to explain many of the classic risk factors originally identified for cervical cancer, such as age at first intercourse and number of lifetime partners [Brinton, 1992]. However, given the evidence from longitudinal cohort studies, that show most HPV infections are transient and only women who harbour a persistent HPV infection are likely to develop a cervical lesion [Moscicki et al., 1993; Hildesheim et al., 1994; Schlecht et al., 2001], other factors are likely to influence the natural history of HPV infections [Bosch et al., 2002].

Viral determinants, including HPV type and viral load, have consistently been identified as important predictors of HPV persistence [Ho et al., 1998; Bosch et al., 2002; Lorincz et al., 2002]. Nonetheless, environmental risk factors, some of which are modifiable, are also thought to play a role in cervical carcinogenesis [Castellsague et al., 2002a]. Recent epidemiological studies have shown that the co-factors most consistently identified in the etiology of cervical neoplasia, after adjusting for the presence of HPV, include high parity and long-term OC use [Lacey et al., 1999; Hildesheim et al., 2001a, Moreno et al., 2002; Munoz et al., 2002], smoking [Szarewski & Cuzick, 1998], and a coinfection with another sexually transmitted disease [Anttila et al., 2001; Smith et al., 2002].

However, it is not clear at which biological stage in HPV carcinogenesis these co-factors are most influential, and results from prospective cohort studies that have attempted to evaluate the effect of hormone contraceptives, diet, alcohol and tobacco consumption on HPV persistence have been inconsistent [Moscicki et al., 1998; Ho et al., 1998; Giuliano et al., 2002]. Furthermore, the definition of HPV persistence has varied significantly between studies, and study designs have differed with respect to varying lengths of follow-up and number of return visits [Hildesheim et al., 1994; Brisson et al., 1996; Moscicki et al., 1998; Ho et al., 1998, Giuliano et al., 2002]. Consequently, comparison of results in the literature can be challenging.

The objectives of this study were to continue the search for epidemiologic determinants of persistent cervical HPV infections. A cohort of young, sexually active women, attending university in Montreal, Canada, was followed over a period of 2 years. This paper presents the results from the analysis of determinants of clearance of type-specific high-risk (HR) and low-risk (LR) HPV infections in this cohort.

METHODS

Subjects

Female students attending either the McGill or the Concordia University Health Clinic were invited to participate if they intended to remain in Montreal, Canada for the next two years and had not required treatment for cervical disease in the last 12 months. Recruitment was initiated in November 1996 and accrual was completed in December 1998.

Procedures

All eligible women were asked to return to the clinic every 6 months over a period of 2 years, for a total of 5 visits. At each visit, a questionnaire was completed and endo- and ectocervical cells from the uterine cervix were collected with two Accelon cervical biosamplers (Medscand Inc., Hollywood, Fla.). A Pap smear was prepared with the first sampler and the remaining cells along with cervical cells collected with a second sampler were used for HPV DNA testing. At enrollment, information from a detailed, self-administered, questionnaire was obtained on potential risk factors such as socio-demographic status, race, diet, smoking history, sexual behaviour, reproductive history, contraceptive and medical history and personal hygiene. An abridged (follow-up) questionnaire designed to measure changes in recent sexual practices and other lifestyle factors was completed at each subsequent visit.

HPV DNA detection

HPV DNA testing by a polymerase chain reaction (PCR) protocol has been described in detail previously [Richardson et al. 2003]. Five µl of DNA purified with QUIAamp columns [Coutlée et al., 1999] (QUIAGEN Inc., CA. USA) was first amplified for β-globin DNA with PC04 and GH20 primers to verify the absence of inhibitors and the integrity of processed DNA [Bauer et al., 1991; Coutlee et al., 2002]. Specimens that were β-globin-positive were further tested with the L1 consensus HPV primers

MY09/MY11 and HMB01 with Amplitag Gold (TaqGold; Perkin-Elmer-Cetus, Norwalk, CT) and the line blot assay (Roche Molecular systems, CA) for the detection of 27 genital HPV genotypes [Gravitt et al., 1998; Coutlee et al., 2002]. HPV types were analyzed individually or in groups according to a classification based on their oncogenic potential. High-risk HPV types included those genotypes that are most frequently found in cervical tumours: HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68. All other individual types that were identified with the line blot assay were classified as low-risk HPV types [Bosch et al., 1995]: 6, 11, 26, 40, 42, 53, 54, 55, 57, 66, 73, 82, 83 and 84 and MM9.

Statistical Methods

Cox's proportional hazards regression [Cox, 1972] was used to estimate the crude and adjusted relative risk of clearance of prevalent and incident HR- or LR-HPV infections over time, according to specific baseline characteristics and behaviours. Separate analyses were carried out for clearance of type-specific, HR- and LR-HPV infections. Variables with repeated measures were represented by their most recent value, using time-dependent covariates [Cox, 1972].

Time to an event was defined as the time between the first HPV positive visit (detected at enrollment or at any follow-up visits) and the first visit when a subject was no longer positive for her longest enduring, type-specific, HR- or LR-infection. Subjects with both HR and LR types at the index visit were assigned to either the high-risk group or the low-risk group, based on the type with the longest duration. If there was a tie, the index visit was assigned to the HR group, in order to make the two groups mutually exclusive. In case of a tie within the high-risk group or low-risk group, all longest persisting infections had to have cleared to be assigned "clearance" status. Those women who completed the study without clearing their infection or were lost to follow-up were censored at their last available visit.

The proportional hazards assumption was tested based on a flexible generalization of the Cox proportional hazards model [Abrahamowicz et al., 1996]. The flexible model allows the hazard ratio for selected exposures to change over time, according to an arbitrary function, the shape of which is estimated from the data using a quadratic regression

spline with 5 degrees of freedom (df). The 4-df likelihood ratio test, comparing the fit of the conventional 1-df proportional hazards model and the flexible 5-df regression spline model was used to verify the null hypothesis that the hazard ratio between those exposed and those not exposed did not change with increasing time since start of infection. All hypotheses were tested at the 0.05 significance level.

We wanted to investigate the role of smoking, alcohol consumption, oral contraceptive use, condom use, and recent STD exposure in clearance of HPV infection. Other variables with at least one category with a p-value less than 0.20, after controlling for age, were also included in the later multivariable models so that they were adjusted for each other. All analyses were performed with SAS version 8.0.

RESULTS

There were 621 women enrolled in the study that completed a questionnaire and consented to HPV DNA testing. The analysis was based on data collected from 2650 completed visits (mean of 4.3 visits/subject) of which 2570 (97.6%) of the cervical specimens were suitable for HPV DNA testing. Women returned every six months, on average (65% returning between 5 and 8 months; range 1-29 months).

A total of 327 (52.7%) women had an HPV infection detected at one or more visits. There were 259 women with a HR-HPV infection and 228 (88%) were positive for a HR-HPV type at one or more visits and had returned for at least one or more visits after HR-HPV detection. Of those 228 women, 37 (16%) had a longer persisting LR-HPV infection and were excluded from the high-risk group, so that 191 (84%) of the women with a high risk HPV infection were included in the analysis for clearance of oncogenic HPV infections. Of those, 91 (48%) women eventually cleared their infection while 100 remained positive at their last visit or were lost to follow-up and, thus, were censored. There were 215 women with a LR-HPV infection and 181 (84%) were positive for a LR-HPV type at one or more visits and had returned for at least one or more visits after LR-HPV detection. Of those 181 women, 94 (52%) had a longer persisting HR infection and were thus excluded from the low-risk group, so that only 87 (48%) of the women with a

low-risk type infection were included in the analysis for clearance of non-oncogenic HPV infections. Of those, 59 (68%) cleared their infection, while 28 women were censored.

The distribution of risk factors (table 7.1) varied significantly between the cohort members from the two participating university clinics. Participants from Concordia University (clinic # 2) had a significantly higher frequency of persistent HR- or LR-HPV, when persistence was defined as 3 or more consecutive visits with the same HR-type or LR-type. The proportional hazards assumption, according to which the hazard ratio is constant over time, was confirmed for all variables of interest (data not shown) except clinic. Thus, given the heterogeneity of the two populations, and the evidence that the effect of clinic on clearance of HPV was not constant over time, all time-to-clearance analyses were stratified by clinic.

There was no consistent relationship between age and time-to HPV clearance (table 7.2). Despite this, all models were adjusted for age since it was considered *a priori* as a potential confounder. Table 7.2 presents the age-adjusted hazard ratios for HR- or LR-HPV clearance according to putative risk factors. The upper part of table 7.2 shows estimates for baseline variables, identified as time-independent variables, while the lower part of table 7.2 shows estimates for variables with repeated measures, identified as time-dependent variables. There was no clear association between sexual activity and clearance of an HPV infection, and recent frequency of sexual intercourse was the only marker of sexual activity that was significantly associated with HPV clearance, although there was not a clear dose-response (table 7.2). Recent and regular use of condoms was significantly associated with faster LR-HPV clearance, only (table 7.2). However, neither the lifetime duration of oral contraceptive use nor the recent use of oral contraceptives was associated with HPV clearance.

When measures of tobacco, alcohol, condom and OC use along with recent STD exposure, and remaining variables with p-values <0.20 in the age-adjusted models, were assessed in a single multivariable model certain effects became stronger (table 7.3). Race was associated with clearance of a HR- but not a LR-HPV infection. Women from

African (Black) descent were significantly more likely to clear their HR-HPV infection than Caucasian women (HR=2.86; 95% CI: 1.18, 6.95). Surprisingly, increased alcohol consumption was associated with a two-fold increased rate of clearance for both LR- and HR-HPV infections, when adjusted for other predictors of clearance, including tobacco consumption. However, women who smoked the equivalent of one or two packs of cigarettes per day for at least one year were less likely to clear their HR-HPV (HR=0.5, 95% CI: 0.2, 1.1) or LR-HPV (HR=0.3, 95% CI: 0.1, 1.1) infections, compared to women who never smoked. Similarly, women that did not have a daily serving of vegetables (carrots, broccoli, cabbage, salad or green beans) were less likely to clear their HR-HPV (HR=0.4, 95% CI: 0.2, 0.7) or LR-HPV (HR=0.5, 95% CI: 0.2, 1.1) infections.

Women with a recent diagnosis of genital warts were four times (HR=4.3, 95% CI: 1.6, 11.5) more likely to clear their LR-HPV infection, when adjusted for other variables, including contraceptive practices. The association between other STDs and HPV persistence could not be evaluated because there were too few women with an HPV infection and a past or recent *Chlamydia trachomatis* or HSV-2 infection. Barrier methods of contraception were associated with clearance of LR-HPV infections, and women who reported using condoms regularly were four times (HR=4.2, 95% CI: 1.7, 10.3) more likely to clear a LR-HPV infection compared to women who did not report using condoms recently (table 7.3). Increased duration of OC use was associated with a faster rate of LR-HPV clearance, but was not statistically significant.

Interestingly, the use of tampons was a strong predictor of HR-HPV persistence only, and women who used tampons, exclusively, were five times (HR=0.2, 95% CI: 0.1, 0.4) less likely to clear their HR-HPV infection compared to women who used sanitary napkins, exclusively.

DISCUSSION

The definition of HPV persistence has varied significantly among studies designed to estimate the average duration of HPV infections and to elucidate the relevant risk factors for persistent HPV infections. Some studies have based HPV persistence on pairs of visits

that were positive for HPV over 3 or more years of follow-up [Ho et al., 1998; Thomas et al., 2000; Ahdieh et al., 2001; Woodman et al., 2001] while other studies have defined persistence based on two visits [Hildesheim et al., 1994; Brisson et al., 1996; Elfgren et al., 2000; Sedjo et al., 2002b; Giuliano et al., 2002b] or on time to HPV clearance [Moscicki et al., 1998]. Studies using Hybrid Capture 2 (Digene Inc., Gaithersburg, MD) (HCII) as the detection system have generally based persistence on oncogenic group since HCII detects a panel of the same 13 oncogenic HPV types described above, rather than individual HPV types. Other groups that have used PCR or Southern blot as their HPV detection system have based persistence on any HPV type or oncogenic group [Brisson et al., 1996; Moscicki et al., 1998; Franco et al., 1999; Woodman et al., 2001; Sedjo et al., 2002a; Giuliano et al., 2002a], or type-specific persistence [Hildesheim et al., 1994; Ho et al., 1998].

Because the majority of studies have not assessed type-specific persistence, the number of persistent cases may have been overestimated by including subjects whose original infection was replaced by a newly acquired infection by a different HPV type. Therefore, women in our cohort were assigned to mutually exclusive viral subgroups, based on their longest type-specific HR- or LR-HPV infection.

Eiological studies on HPV persistence have varied widely with respect to length of time between return visits. This difference in time can affect the results as women with a longer time interval between two visits (>12 months) are significantly less likely to have a persistent HPV infection [Hildesheim et al., 2001a]. Extreme variation in length of time between visits could also have biased our estimated time to HPV clearance. While the majority of women in our cohort returned within 6 months of each visit, there were a small proportion of women whose time interval between visits was greater than one year. If this subgroup of women had differentially cleared their infection at a much earlier date prior to their return visit date, the average time to clearance could be inflated. However, the association between putative risk factors and clearance would only be distorted if certain exposure-outcome combinations were associated differentially with time between

visits. Given that the participants were unaware of their HPV status until the end of the study, this is an unlikely scenario.

Rather than study predictors of persistence, based on a pre-specified number (or pairs) of HPV positive visits, we decided to try and identify determinants of HPV clearance, using survival analysis techniques. An advantage of using survival analysis is that it can model the natural history of a type-specific HPV episode, and does not rely on arbitrary definitions of persistent and transient infections. Moreover, the effect of time-dependent repeated measures on HPV clearance could be assessed using an extension of the Cox Proportional Hazards model. This model could accommodate for changes in behaviour over time (when available), and allow us to detect potential effects of recent changes in risk behaviour that might impact on the rate of clearance. Nonetheless, for certain variables such as tobacco and OC use, we felt it was more biologically relevant in the etiology of HPV persistence to include the measures of "cumulative exposure" in the final multivariable models, rather then the measures of recent exposure.

Both prevalent and incident infections were included in our analysis of clearance, because of the limited number of incident HPV infections in our dataset. One concern with combining prevalent and incident HPV data is that the prevalent HPV infections are left censored, and the time at which a woman is first HPV positive cannot be easily approximated. However, the median duration of a prevalent type-specific HPV infection was only about 2 months longer than the median duration of an incident type-specific HPV infection in our study. When we evaluated the association between selected risk factors and clearance of incident HPV infections the hazard ratios did not differ materially from the point estimates presented in tables 7.2 and 7.3, although the confidence intervals were wider (data not shown).

The viral-host interaction is believed to play a crucial role in HPV persistence. Furthermore, functional differences between oncoproteins of viral HPV subgroups are believed to explain the superior ability of HR-HPV types to cause cellular transformation, proliferation and lesion progression [de Sanjose et al., 1994; Ho et al., 1995; zur Hausen,

2000]. Thus, if there are biological distinctions between the molecular mechanisms of LR- and HR-HPV induced infections, there may also be distinct sets of risk factors for persistence associated with these HPV subgroups.

However, the overall profile of predictors for LR- and HR-HPV clearance was quite similar in our cohort. We failed to observe a significant association between age and HPV clearance even though persistence of HPV infections has been significantly associated with older age (>30) in some [Ho et al., 1998] [Hildesheim et al., 1994], but not all studies [Brisson et al., 1996]. However, the vast majority (>90%) of women in our cohort were younger than 30 years of age. Recent sexual activity has not been a significant predictor of HPV persistence in most cohort studies [Hildesheim et al., 1994] [Moscicki et al., 1998; Ho et al., 1998], although number of lifetime partners has been associated with an elevated risk for HPV persistence in some studies [Brisson et al., 1996; Elfgren et al., 2000]. Markers of sexual activity, apart from recent frequency of sex, explained very little about HPV clearance in our cohort and, the absence of a dose-response for frequency of sex limits the interpretation of this association.

We observed that women who had smoked one or two packs of cigarettes per day for at least one year were nearly half as likely to clear their HPV infections then never smokers, although the association was marginally non-significant, and there was not a clear dose-response. The association between smoking tobacco and HPV persistence has not been consistently demonstrated in the literature. Non-smokers were surprisingly more likely to have a persistent HPV infection than current smokers among a cohort of university students in New York, followed every 6 months for over two years [Ho et al., 1998] and middle-aged women in Oregon, who returned for one follow-up visit [Hildesheim et al., 1994]. Conversely, Giuliano and collaborators [Giuliano et al., 2002] observed that smokers were less likely to clear an HPV infection than non-smokers, among a cohort of women that returned for 2 follow-up visits over an average of ten months. Others have not observed any association between tobacco use and HPV persistence [Brisson et al., 1996] or clearance [Moscicki et al., 1998]. However, only one of these studies [Giuliano et al., 2002] evaluated the cumulative intensity of tobacco exposure. When we evaluated

the effect of recent number of cigarettes smoked or smoking status (never, former, current) in the final multivariable model neither variable was associated with HPV clearance (data not shown).

Despite the absence of a dose-response with increasing duration and intensity of tobacco exposure in our cohort, there are still a few speculative biological mechanisms that could explain the association between tobacco exposure and HPV persistence. The most popular theory posits that tobacco exposure affects the host immune-surveillance system against viral infections, as demonstrated by reduced markers of immune function *in vitro* [Poppe et al., 1995; Castelsague et al., 2002]. There is also preliminary laboratory evidence to suggest that certain HPV infected cells may be more susceptible to DNA damage from specific tobacco carcinogens [Melikian et al., 1999]. Nonetheless, based on the current data in the literature, it is still difficult to conclude at which stage in the natural history of HPV cervical carcinogenesis, tobacco exposure would exert most influence.

Other factors that may reduce immune function or impact on the health of the cervical squamous epithelial environment could include excessive alcohol consumption [Reichman et al., 1993; Hankinson et al., 1995] and dietary patterns deficient in essential micronutrients [Sedjo et al., 2002; Sedjo et al., 2002b]. However, our data did not support previous observations that showed increased alcohol consumption was either associated with increased HPV persistence [Ho et al., 1998], or not associated with HPV clearance at all [Moscicki et al., 2001]. We observed that women who drank more than 3 alcoholic beverages per week were more likely to clear their HPV infections, compared to women who did not drink alcohol when adjusted for potential confounders such as tobacco and condom use. However, approximately 90% of the cohort members in our study drank less than 1 drink per day. Therefore, the effect of heavy drinking on HPV clearance could not be fully assessed. While it is tempting to speculate that moderate alcohol consumption (~1 drink/day) may be associated with lower levels of anxiety and stress that may, in turn, have beneficial effects on boosting immunity, there is no experimental data in the HPV literature, to currently support this claim.

A limited food frequency questionnaire was completed by our cohort members at enrollment only, and the main goal was to characterize general eating patterns with respect to dairy products, and fruits and vegetables that were high in vitamins C & E, caretenoids and folate. Nonetheless, despite our crude measurement of dietary patterns, we observed that women that consumed one or more servings of vegetables per day cleared their HPV infections significantly more quickly then women who did not consume vegetables daily. Our results add to the mounting evidence that fresh vegetables may play a protective role against persistent HPV infections. A recent study that measured food intake with a food frequency questionnaire and also measured levels of specific micronutrients in blood samples, observed that a higher level of vegetable consumption was associated with a 54% decrease risk of HPV persistence in a cohort of young women from Arizona [Sedjo et al., 2002b]. A dose response was also observed between increasing vegetable fibre in the diet and reduced risk of HSIL, after adjusting for HPV status, in a population-based case-control study in Sweden [Kjellberg et al., 2000]. Two different mechanistic pathways have been proposed to support the protective effects of a healthy diet against cervical carcinogenesis. Anti-oxidant nutrients such as vitamins C & E, lycopene and caretenoids may act early in cervical carcinogenesis by helping to prevent cellular and DNA damage from reactive oxygen species, caused from exposure to tobacco or cervical inflammation [Castle and Giuliano 2003]. Folate, as well as vitamins B_{12} and B_6 and methionine may also play a role in decreasing viral proliferation through their role in DNA methylation [Sedjo et al., 2002]. Future studies might want to focus on specific micronutrients and attempt to identify the pathway most likely involved in early HPV carcinogenesis (persistence).

Hygienic practices have frequently been proposed as explanatory variables for observed differences in HPV prevalence, but associations have been inconsistent [Ho et al., 1998; Richardson et al., 2000; Rousseau et al., 2000]. Our data showed women that used tampons instead of sanitary napkins, during menstruation, were significantly less likely to clear their HR-HPV infection. One cohort study observed an elevated risk for cumulative HR-HPV among women who used tampons instead of sanitary napkins during

menstruation [Rousseau et al., 2000]. While the mechanism is currently unknown, it is possible that tampons may serve to spread the viral infection to different locations in the cervix, thereby increasing the area and possibly the severity of the HPV infection.

Recent regular condom use was protective against LR-HPV persistence in our study. While condom use and other barrier contraceptives have not been shown to be associated with HPV persistence in previous studies [Brisson et al., 1996; Moscicki et al., 1998], regular use of condoms has been associated with a reduction in risk for high-grade squamous intraepithelial lesions among HPV positive women in two separate cohorts [Ho et al., 1998; Hildesheim et al., 2001a]. These results are fairly unexpected, since it has generally been thought that if condoms do confer protection, it would be further upstream, by preventing transmission of HPV, not lesion development. However, it is possible that, if condoms do protect against acquisition of a new HPV infection, then extended persistent HPV episodes (re-infection with the same type) may be prevented. Another possibility is that while condoms may not entirely protect against transmission, they may help reduce the viral load of HPV or protect against other STDs, that may interact with HPV and worsen the natural course of infection [Hildesheim et al., 2001a].

Surprisingly, a recent history of genital warts was significantly associated with an increased likelihood of clearance of a LR-HPV infection in our cohort. These results are in contrast to results from another cohort study in San Francisco, that showed young women were half as likely to clear either a HR-or a LR-HPV infection [Moscicki et al., 1998], if they had a recent history of vulvar condylomas. The discrepant results may in part be explained by differences between the two cohorts. The San Francisco cohort members were younger (13-21 years old) on average, and considered a "high-risk" population since they were recruited from family-planning clinics rather then a university setting. It is possible that unknown factors associated with cohort membership may have modified the association between genital warts and HPV clearance observed in the two studies. It is also conceivable that newer treatment for genital warts with anti-viral properties such as Imiquimod [Ferenczy & Franco, 2002], could have exerted the

seemingly protective effect we observed in our cohort. Unfortunately, we failed to collect information on treatment for STDs, which prevented us from examining this possibility.

In conclusion, viral characteristics (i.e. type and load) and the host environment (i.e. genetic susceptibility, immunosuppression) are currently the most important determinants of HPV persistence [Ahdieh et al., 2001; Bosch et al., 2002; Clarke & Chetty, 2002]. Therefore, until a vaccine or anti-viral medication is available to treat HPV infections, there appears to be little that a physician can do for a HPV positive patient, other than see her regularly for cervical cytology screening. However, our results suggest that there may be some proactive measures that a woman can take to modify her risk for HPV persistence. The cessation of tobacco use, increased vegetable consumption and use of sanitary napkins instead of tampons during the course of an HPV infection are actions that may potentially lead to a faster rate of HPV clearance.

Acknowledgments
This study was funded by a grant from the Canadian Institutes of Health Research (CIHR) (MT-13649). HR is recipient of a Pre-doctoral Scholarship, MA of a CIHR Scientist Award, and EF of a Distinguished Scientist Award, all from the CIHR. FC is recipient of a Chercheur Boursier National from the Fonds de la Recherche en Santé du Québec. The authors are indebted to Suzanne Dumais and the staff at the McGill and Concordia Student Health Services who helped with subject accrual and follow-up. The authors also thank Roche Molecular Systems who generously supplied us with all the reagents necessary for HPV genotyping.

Table 7.1
Distribution of baseline characteristics between participants from the two participating university health clinics

Baseline characteristics	Total Cohort N=621 n (%)	University Clinic #1 N=421 %	University Clinic #2 N=200 %
Age			
17-20	187 (30.1)	41.3	6.5
21-23	228 (36.7)	34.4	41.5
24-26	102 (16.4)	11.2	27.5
27+	104 (16.7)	13.1	24.5
		P-value (LR- $\chi 2$) =	0.000
Race			
White	513 (82.6)	82.6	82.5
Asian	62 (10.0)	11.9	6.0
Black	29 (4.7)	3.6	7.0
Hispanic	17 (2.7)	1.9	4.5
		P-value (LR- $\chi 2$) =	0.009
Smoking status			
Never	374 (60.2)	67.2	45.5
Former	98 (15.8)	13.1	21.5
Current	149 (24.0)	19.7	33.0
		P-value (LR- $\chi 2$) =	0.000
Cigarette consumption:			
None	411 (66.2)	71.5	48.0
1-5/day	119 (19.2)	13.1	22.0
>5/day	91 (14.7)	15.4	30.0
		P-value (LR- $\chi 2$) =	0.000
Pack-years smoked			
Non-smoker	378 (60.9)	67.7	46.5
<1 pack-year	101 (16.3)	16.6	15.5
1-2 pack-years	58 (9.3)	7.1	14.0
>2 pack-years	84 (13.5)	8.6	24.0
		P-value (LR- $\chi 2$) =	0.000
Alcohol consumption			
0 drinks/wk	226 (36.4)	39.2	30.5
1-3 drinks/wk	179 (28.8)	29.2	28.0
>3 drinks/wk	216 (34.8)	31.6	41.5
		P-value (LR- $\chi 2$) =	0.035
Age at first sex encounter			
19+ years	174 (28.0)	29.2	25.5
16-18 yeas	314 (50.6)	52.5	46.5
<16 years	133 (21.4)	18.3	28.0
		P-value (LR- $\chi 2$) =	0.025

Table 7.1 (continued)
Distribution of baseline characteristics between participants from the two participating university health clinics

Baseline characteristics	Total Cohort N=621 n (%)	University Clinic #1 N=421 %	University Clinic #2 N=200 %
No. lifetime sex partners			
1 lifetime partners	147 (23.7)	27.6	15.5
2-4 lifetime partners	196 (31.6)	34.2	26.0
4-9 lifetime partners	161 (25.9)	25.7	26.5
10+ lifetime partners	117 (18.8)	12.6	32.0
		P-value (LR- $\chi 2$) =	0.000
No. recent partners			
0 partners in 6 mo.	39 (6.3)	5.9	7.0
1 partner in 6 mo.	345 (55.6)	55.3	56.0
2+ partners in 6 mo.	237 (38.1)	38.7	37.0
		P-value (LR- $\chi 2$) =	0.837
New no. sex partners			
0 new partners in 6 mo.	260 (41.9)	39.2	47.5
1 new partner in 6 mo.	202 (32.5)	34.9	27.5
2+ new partners in 6 mo.	159 (25.6)	25.9	25.0
		P-value (LR- $\chi 2$) =	0.101
Frequency of vaginal sex			
<1/wk	111 (17.9)	19.2	15.0
1-2/wk	266 (42.8)	41.3	46.0
>2/wk	244 (39.3)	39.4	39.0
		P-value (LR- $\chi 2$) =	0.351
Oral contraceptive (OC) use:			
Never	145 (26.3)	25.7	18.5
Sometimes	64 (10.3)	8.6	14.0
Always	412 (66.3)	65.8	67.5
		P-value (LR- $\chi 2$) =	0.031
Duration of OC use			
(0 years)	145 (23.3)	25.7	18.5
<1year	121 (19.5)	21.6	15.0
1-5 years	252 (40.6)	40.1	41.5
5+ years	103 (16.6)	12.6	25.0
		P-value (LR- $\chi 2$) =	0.000
Condom use:			
Never	54 (8.7)	8.1	10.0
Sometimes	208 (33.5)	30.9	39.0
Always	359 (57.8)	61.0	51.0
		P-value (LR- $\chi 2$) =	0.061

Table 7.1 (continued)
Distribution of baseline characteristics between participants from the two participating university health clinics

Baseline characteristics	Total Cohort N=621 n (%)	University Clinic #1 N=421 %	University Clinic #2 N=200 %
Menstrual products			
Pads	107 (17.2)	14.7	22.5
Tampons	135 (21.7)	21.9	21.5
Pads & tampons	379 (61.0)	63.4	56.0
		P-value (LR- $\chi 2$) =	0.055
Wash after sex			
Never /rarely	208 (33.5)	35.2	30.0
Sometimes	361 (58.1)	57.2	60.0
Always	52 (8.4)	7.6	10.0
		P-value (LR- $\chi 2$) =	0.335
Vegetable consumption			
1+/day	132 (21.3)	22.8	18.0
<1/day	489 (78.7)	77.2	82.0
		P-value (LR- $\chi 2$) =	0.172
History of *Chlamydia*			
Never	589 (94.8)	97.1	90.0
Ever	32 (5.2)	2.9	10.0
		P-value (LR- $\chi 2$) =	0.000
History of Warts			
Never	537 (86.5)	89.8	79.5
Ever	84 (13.5)	10.2	20.5
		P-value (LR- $\chi 2$) =	0.000
Cumulative HPV status:			
HPV negative (at all visits)	294 (47.3)	53.4	34.5
HPV positive (at 1+ visits)	327 (52.7)	46.6	65.5
		P-value (LR- $\chi 2$) =	0.000
Persistent HR-HPV			
Negative or transient	548 (88.2)	90.5	83.5
> 3 consecutive visits HR-HPV+	73 (11.8)	9.5	16.5
		P-value (LR- $\chi 2$) =	0.011
Persistent LR-HPV			
Negative or transient	583 (93.9)	95.5	90.5
> 3 consecutive visits LR-HPV+	38 (6.1)	4.5	9.5
		P-value (LR- $\chi 2$) =	0.015
Cumulative SIL status:			
Normal cytology (at all visits)	575 (92.6)	91.9	94.0
SIL (at 1+ visits)	46 (7.4)	8.1	6.0
		P-value (LR- $\chi 2$) =	0.000

Table 7.2
Determinants of clearance of type-specific high-risk and low-risk HPV infections: crude & adjusted for age

	High oncogenic risk (n=191)		Low oncogenic risk (n=87)	
	HR[a] (95%CI[b]) Crude	HR (95%CI) Age adjusted	HR (95%CI) Crude	HR (95%CI) Age adjusted
Time-independent variables				
Age				
17-20 years	REF		REF	
21-23 years	1.1 (0.7, 1.7)		0.9 (0.5, 1.7)	
24-26 years	0.6 (0.3, 1.1)		0.8 (0.4, 1.6)	
27+ years	0.8 (0.4, 1.7)		0.9 (0.4, 1.8)	
Age at first sex encounter				
19+ years	REF	REF	REF	REF
16-18 years	0.8 (0.5, 1.4)	0.8 (0.5, 1.3)	1.3 (0.7, 2.5)	1.2 (0.6, 2.1)
<16 years	0.9 (0.5, 1.6)	0.8 (0.5, 1.5)	1.0 (0.5, 2.0)	0.9 (0.4, 1.9)
No. lifetime sex partners				
1 sex partner	REF	REF	REF	**REF***
2-4 sex partners	0.9 (0.4, 1.7)	0.8 (0.4, 1.6)	0.5 (0.2, 1.1)	**0.4 (0.2, 1.0)**
5-9 sex partners	0.6 (0.3, 1.2)	0.6 (0.3, 1.3)	0.8 (0.4, 1.8)	**0.8 (0.4, 1.8)**
10+ sex partners	0.7 (0.4, 1.5)	0.8 (0.4, 1.6)	0.5 (0.2, 1.2)	**0.5 (0.2, 1.2)**
New no. sex partners				
0 new partners	REF	**REF***	REF	REF
1 new partner	1.6 (0.9, 2.6)	**1.5 (0.9, 2.5)**	0.7 (0.4, 1.3)	0.7 (0.4, 1.3)
2+ new partners	1.5 (0.9, 2.4)	**1.4 (0.8, 2.3)**	0.8 (0.4, 1.4)	0.7 (0.4, 1.3)
Race				
White	REF	**REF***	REF	REF
Black	1.7 (0.8, 3.5)	**1.8 (0.8, 3.7)**	0.6 (0.2, 2.0)	0.7 (0.2, 2.2)
Hispanic	1.8 (0.6, 5.8)	**1.9 (0.6, 6.0)**	2.6 (0.4, 19.3)	3.0 (0.4, 22.3)
Asian	1.1 (0.5, 2.4)	**1.2 (0.5, 6.0)**	0.8 (0.3, 2.1)	1.0 (0.3, 2.7)

*[a]HR=Hazard ratio; [b]CI=Confidence Interval (*p<0.20; **p<0.05)*

Table 7.2 (continued)
Determinants of clearance of type-specific high-risk and low-risk HPV infections: crude & adjusted for age

	High oncogenic risk (n=191)		Low oncogenic risk (n=87)	
	HR[a] (95%CI[b]) Crude	HR (95%CI) Adjusted	HR (95%CI) Crude	HR (95%CI) Adjusted
Duration of OC use				
0 years	REF	**REF***	REF	REF
<1year	1.4 (0.8, 2.5)	**1.4 (0.8, 2.5)**	0.7 (0.2, 1.9)	0.7 (0.2, 2.0)
1-5 years	0.8 (0.5,1.4)	**0.8 (0.5, 1.4)**	1.0 (0.5, 1.4)	1.0 (0.5, 2.0)
5+ years	0.7 (0.4, 1.4)	**0.8 (0.4, 1.7)**	1.2 (0.6, 2.6)	1.4 (0.6, 3.1)
Menstrual products				
Pads	REF	**REF****	REF	REF
Tampons	0.3 (0.2, 0.6)	**0.3 (0.1, 0.6)**	1.0 (0.5, 1.9)	1.0 (0.5, 2.0)
Pads & tampons	0.6 (0.3, 0.9)	**0.5 (0.3, 0.9)**	0.7 (0.4, 1.4)	0.8 (0.4, 1.5)
Vegetable consumption				
1+/day	REF	**REF****	REF	**REF***
<1/day	0.5 (0.3, 0.8)	**0.6 (0.3, 0.9)**	0.6 (0.4, 1.1)	**0.6 (0.4, 1.1)**
Time-dependent variables				
Smoking status				
Never	REF	REF	REF	REF
Former	1.0 (0.5, 2.1)	1.0 (0.5, 2.0)	0.7 (0.3, 1.9)	0.7 (0.3, 1.8)
Current	0.8 (0.5, 1.2)	0.7 (0.5, 1.2)	1.2 (0.7, 2.1)	1.1 (0.6, 2.1)
No. cigarettes/day				
Non-smoker	REF	REF	REF	Ref
<1/day	0.4 (0.2, 1.1)	0.4 (0.2, 1.0)	1.3 (0.6, 2.6)	1.2 (0.6, 2.6)
1-5/day	1.0 (0.6, 1.9)	0.9 (0.5, 1.8)	1.2 (0.6, 2.5)	1.1 (0.5, 2.5)
>5/day	0.9 (0.5, 1.6)	0.8 (0.4, 1.5)	1.4 (0.7, 2.9)	1.4 (0.6, 2.9)

*[a]HR=Hazard ratio; [b]CI=Confidence Interval (*p<0.20; **p<0.05)*

Table 7.2 (continued)
Determinants of clearance of type-specific high-risk and low-risk HPV infections: crude & adjusted for age

	High oncogenic risk (n=191)		Low oncogenic risk (n=87)	
	HR[a] (95%CI[b]) Crude	HR (95%CI) Adjusted	HR (95%CI) Crude	HR (95%CI) Adjusted
Pack-years smoked				
Non-smoker	REF	**REF***	REF	REF
<1 pack-years	0.9 (0.5, 1.6)	**0.8 (0.5, 1.5)**	1.4 (0.7, 2.6)	1.3 (0.7, 2.5)
1-2 pack-years	0.6 (0.3, 1.3)	**0.6 (0.3, 1.2)**	0.6 (0.2, 1.9)	0.5 (0.2, 1.9)
>2 pack-years	0.7 (0.4, 1.4)	**0.7 (0.4, 1.4)**	1.1 (0.5, 2.2)	1.1 (0.5, 2.2)
Alcohol consumption				
0 drinks/wk	REF	REF	REF	REF
1-3 drinks/wk	1.1 (0.6, 2.0)	1.1 (0.6, 1.9)	1.4 (0.7, 2.9)	1.4 (0.6, 3.2)
>3 drinks/wk	1.0 (0.6, 1.9)	1.0 (0.5, 1.8)	1.9 (0.9, 3.9)	1.9 (0.9, 4.3)
Oral contraceptive use				
Never	REF	REF	REF	REF
Sometimes	0.9 (0.3, 2.4)	0.8 (0.3, 2.2)	0.0 (0.0, -)	0.0 (0.0, -)
Always	0.9 (0.6, 1.4)	0.8 (0.3, 1.3)	1.1 (0.6, 1.9)	1.1 (0.6, 1.9)
Condom use				
Never	REF	REF	REF	**REF****
Sometimes	0.9 (0.5, 1.5)	0.8 (0.5, 1.5)	1.2 (0.6, 2.5)	**1.2 (0.5, 2.5)**
Always	1.0 (0.6, 1.7)	1.1 (0.6, 1.7)	2.3 (1.2, 4.6)	**2.3 (1.1, 4.7)**
Foam use				
Never	REF	REF	REF	REF
Sometimes	0.8 (0.2, 2.5)	0.8 (0.3, 2.7)	0.4 (0.1, 1.8)	0.4 (0.1, 1.6)
Always	1.7 (0.4, 7.1)	1.6 (0.4, 6.8)	1.9 (0.4, 8.7)	1.9 (0.4, 8.7)
Wash after sex				
Never /rarely	REF	REF	REF	REF
Sometimes	1.2 (0.8, 1.9)	1.1 (0.7, 1.8)	0.9 (0.5, 1.5)	0.8 (0.5, 1.4)
Always	0.8 (0.4, 1.6)	0.8 (0.4, 1.6)	0.8 (0.3, 2.0)	0.7 (0.3, 1.9)

*[a]HR=Hazard ratio; [b]CI=Confidence Interval (*p<0.20; **p<0.05)*

Table 7.2 (continued)
Determinants of clearance of type-specific high-risk and low-risk HPV infections: crude & adjusted for age

	High oncogenic risk (n=191)		Low oncogenic risk (n=87)	
	HR[a] (95%CI[b]) Crude	HR (95%CI) Adjusted	HR (95%CI) Crude	HR (95%CI) Adjusted
Frequency of sex				
<1/wk	REF	**REF****	REF	REF
1-2/wk	0.5 (0.3, 0.9)	**0.5 (0.3, 0.9)**	1.5 (0.7, 2.9)	1.5 (0.7, 3.0)
3+/wk	0.9 (0.5, 1.4)	**0.8 (0.5, 1.3)**	1.3 (0.7, 2.6)	1.3 (0.7, 2.5)
Anal sex				
Never since last visit	REF	**REF***	REF	REF
Ever since last visit	1.5 (0.9, 2.5)	**1.5 (0.9, 2.5)**	0.9 (0.4, 1.6)	0.9 (0.4, 1.7)
History of Chlamydia			Not possible to calculate	Not possible to calculate
No	REF	REF		
Yes	1.4 (0.3, 6.1)	1.4 (0.3, 5.9)		
History of warts				
No	REF	REF	REF	**REF***
Yes	0.5 (0.1, 2.1)	0.5 (0.1, 2.0)	2.0 (0.9, 4.3)	**2.0 (0.9, 4.3)**
Vaginal irritation				
Never	REF	**REF***	REF	REF
1-2/since last visit	1.6 (1.0, 2.5)	**1.5 (0.9, 2.4)**	1.1 (0.6, 2.0)	1.1 (0.6, 2.0)
>2/ since last visit	1.6 (0.9, 3.1)	**1.6 (0.9, 3.1)**	1.1 (0.4, 2.7)	1.1 (0.4, 2.6)

*[a]HR=Hazard ratio; [b]CI=Confidence Interval (*p<0.20; **p<0.05)*

Table 7.3
Final multivariable model for clearance of type-specific high-risk and low-risk HPV infections

Variables	High -risk HPV clearance (n=191)		Low -risk HPV clearance (n=87)	
	Hazard Ratio	(95% CI)	Hazard Ratio	(95% CI)
Age				
17-20 years	REF	REF	REF	REF
21-23 years	1.3	(0.7, 2.5)	1.4	(0.5, 3.9)
24-26 years	0.7	(0.3, 1.6)	0.9	(0.3, 2.8)
27+ years	0.7	(0.3, 1.9)	1.3	(0.4, 4.6)
Race			not included in final	
White	REF	REF	model*	
Black	3.5	(1.2, 9.4)		
Hispanic	3.7	(0.6, 22.3)		
Asian	1.5	(0.6, 4.0)		
Duration of OC use				
0 years	REF	REF	REF	REF
<1year	1.7	(0.8, 3.5)	1.2	(0.3, 4.2)
1-5 years	1.2	(0.6, 2.4)	1.9	(0.7, 5.5)
5+ years	1.0	(0.4, 2.5)	2.3	(0.6, 9.2)
Vegetable consumption				
(1+/day)	REF	REF	REF	REF
<1/day	0.4	(0.2, 0.7)	0.5	(0.2, 1.1)
Sanitary products			not included in final	
Pads only	REF	REF	model*	
Tampons	0.2	(0.1, 0.4)		
Tampons & pads	0.4	(0.2, 0.7)		

** These variables were not included in the final model for clearance of low-risk infections because of a clear lack of association in earlier models.*
All variables were adjusted for each other in the model.

Table 7.3 (continued)
Final multivariable model for clearance of type-specific high-risk and low-risk HPV infections

Variables	High -risk HPV clearance (n=191)		Low -risk HPV clearance (n=87)	
	Hazard Ratio	95% C.I.	Hazard Ratio	95% C.I.
Pack-years smoked				
None	REF.	REF	REF	REF
<1 pack-year	0.7	(0.4, 1.4)	0.8	(0.3, 1.8)
1-2 pack-years	0.5	(0.2, 1.1)	0.3	(0.1, 1.1)
>2 pack-years	0.6	0.3, 1.3)	0.7	(0.3, 1.9)
Alcohol consumption				
0 drinks/wk	REF	REF	REF	REF
1-3 drinks/wk	2.1	(1.0, 4.5)	1.1	(0.4, 2.7)
>3 drinks/wk	2.0	(1.0, 4.9)	1.9	(0.7, 5.3)
Condom use				
Never	REF	REF	REF	REF
Sometimes	0.8	(0.4, 1.5)	2.3	(0.9, 5.8)
Always	1.0	(0.6, 1.7)	4.2	(1.7, 10.3)
Frequency of sex				
<1/wk	REF	REF	not included in final	
1-2/wk	0.4	(0.2, 0.7)	model*	
3+/wk	1.0	(0.6, 1.8)		
Warts				
No	REF	REF	REF	REF
Yes	0.4	(0.1, 1.9)	4.3	(1.6, 11.5)
Vaginal irritation				
Never	REF	REF	not included in final	
1-2 since last visit	1.6	(0.9, 2.8)	model*	
>2/ since last visit	1.9	(0.9, 3.9)		

* *These variables were not included in the final model for clearance of low-risk infections because of a clear lack of association in earlier models.*
All variables were adjusted for each other in the model.

CHAPTER 8: MANUSCRIPT IV

Infection with HPV 16 or HPV 18 variants and risk of low-grade squamous intraepithelial lesions in young women

8.1 Preamble

Worldwide, HPV-16 and HPV-18 are two of the most common HPV infections found in women with normal cytology and in cases with invasive cervical cancer. However, very few women with HPV 16 or -18 infections actually go on to develop cervical cancer, in part because of screening programs and possibly because of other modifiable co-factors, as discussed in manuscript III. Nonetheless, HPV viral characteristics are also thought to play an important role in the transition from a transient HPV infection to HPV persistence or to cervical neoplasia and cancer. The relationship of HPV-16 and -18 polymorphic variations (variants) and cervical lesions has been addressed in a few studies in an attempt to help explain the progression of an HPV infection to a high-grade squamous intraepithelial lesion or cancer [Zehbe et al., 1998; Maciag et al., 2000; Berumen et al., 2001; Hildesheim et al., 2001b; Xi et al., 2002].

The objectives of manuscript IV were to describe the distribution of HPV-16 and HPV-18 variants in a young, healthy population and study the effect of HPV variants on HPV persistence and risk of incident low-grade squamous intraepithelial lesions (LSIL).

The previous three manuscripts addressed the dynamics and determinants of HPV infections prior to the development of cytological abnormalities. Manuscript IV focuses on a slightly later event in the natural history of HPV, subsequent to a persistent HPV infection and aims to identify which HPV-16 and -18 variants were most prevalent in incident LSILs.

Infection with HPV 16 or HPV 18 variants and risk of low-grade squamous intraepithelial lesions in young women

Harriet Richardson[1,2], François Coutlée[5], Michal Abrahamowicz[2], Pierre Tellier[3], Gail Kelsall[1,3], Hélène Voyer[5], Alex Ferenczy[1,4], and Eduardo L. Franco[1,2]

Affiliation of authors: (1) Departments of Oncology, (2) Epidemiology & Biostatistics, (3) Family Medicine, and (4) Pathology, McGill University, Montreal, Canada; (5) Laboratoire de Virologie Moleculaire, Centre de Recherche et Departement de Microbiologie et Infectiologie, Hôpital Notre-Dame du Centre Hospitalier de l'Université de Montréal, Montreal, Canada.

Correspondence to: Prof. EL. Franco, Department of Oncology, McGill University, 546 Pine Avenue West Montreal, QC, Canada H2W 1S6 (e-mail: eduardo.franco@mcgill.ca).

Abstract

Background: Recent reports have described substantially elevated risks for high-grade cervical neoplasia and carcinoma associated with some molecular variants of HPV 16 and 18. The objectives of this study were to evaluate the association between infections with European (E) or Non-European (NE) variants of HPV 16 or HPV 18, and newly acquired low-grade squamous intraepithelial lesions (LSIL), in a cohort of university students.

Methods: Female university students (n=621) in Montreal were followed for 24 months at 6-month intervals. At each visit a cervical cell specimen was collected. HPV DNA was detected using the MY09/MY11 PCR protocol and 27 HPV genotypes were identified by the line blot assay (Roche Molecular Systems Inc.). Clinical samples that were positive for HPV types 16 or 18 were analyzed for HPV 16 and 18 variants. A PCR-sequencing method was used with specific HPV-16 and HPV-18 primers designed to flank nucleotide positions 7478-7841 and 7485-7805, respectively, in the viral genome's long control region. Variants were classified into two groups: European (E) or Non-European (NE). Logistic regression was used to evaluate the association between an incident LSIL infection and HPV status. The Kaplan-Meier technique was used to estimate the cumulative probability of acquiring or clearing an LSIL according to LR- or HR-HPV types or HPV16/18 variant status.

Results: The mean duration of incident HPV16/18 E and NE variant infections was 11.3 months and 7.0 months, respectively. The overall incidence rate for LSIL among women in the cohort was 2.4 per 100 person-years, and was 4.1 per 100 person-years among women with HPV. Women with an NE variant (OR=38.1 95% CI: 4.7, 310.9) or an E variant (OR=10.6 95% CI: 2.1, 52.4) infection were significantly more likely to have an incident LSIL compared to women who were HPV negative.

Conclusion: This study suggests that intratype variability in HPV 16 or -18 may mediate LSIL development and Non-European variants may be more strongly associated with incident LSIL than European variants.

INTRODUCTION

A persistent HPV infection, especially with HPV 16 or -18, is considered one of the strongest predictors of cervical intraepithelial neoplasia (CIN) or cancer [Ho et al., 1995; Moscicki et al., 1998; Wallin et al., 1999; Schlecht et al., 2001], and clearance of HR-HPV infections is significantly associated with subsequent regression of low-grade cervical lesions [Nobbenhuis et al., 2001].

The identification of molecular variants of HPV 16 or -18 was originally of interest to researchers mapping the evolution of different HPV types geographically [Ong et al., 1993; Ho et al., 1993]. Studies of the noncoding regions of the HPV 16 genome (long control region or LCR) and the E6 gene revealed that recombination between variants was very rare [Ho et al., 1993; Yamada et al., 1995]. Each HPV genotype can further be classified into variants defined by a limited number of genomic variations in coding and non-coding regions [Ho, 1991]. The heterogeneity of HPV-16 genes demonstrates a strong intergene sequence covariation [Chan et al., 1992]. HPV 16 evolved along 5 major geographic lineages and, for simplicity purposes, can be classified into two variant groups, namely European (E) or Prototype-Like (PL) variants (that bear strong homology to the prototype isolated in Europe) and the Non-European (NE) or Non-Prototype-Like (NPL) variants (originating predominantly in Africa, Asia or India) [Ong et al., 1993; Ho et al., 1993; Villa et al., 2000; Xi et al., 2002].

Some longitudinal cohort studies incorporated the application of DNA sequencing techniques into their study design to detect molecular variants of specific HPV types [Villa et al., 2000; Hildesheim et al., 2001b; Xi et al., 2002]. Viral typing was useful for screening transient infections from persistent infections, but viral typing could not distinguish between two consecutive transient infections with the same HPV type and a persistent infection caused by the same variant. Molecular variant analysis was thought to allow for interpretation of persistence on firmer grounds [Franco et al., 1994].

Subsequent reports from prospective cohort studies have indicated that individual cases of persistently detected HPV 16 or -18 infections are usually of the same molecular

variant [Xi et al., 1995; Villa et al., 2000]. However, results about differential duration (persistence) of different HPV types including infections with HPV 16 variants have been inconsistent [Londesborough et al., 1996; Villa et al., 2000; Xi et al., 2002]. Moreover, some recent cohort and cross-sectional studies have observed a greater risk of high-grade cervical neoplasia associated with certain HPV 16 and HPV 18 variants as compared with other variants [Zehbe et al., 1998; Villa et al., 2000; Berumen et al., 2001; Hildesheim et al., 2001b; Xi et al., 2002]. Consequently, it has been proposed that variants with greater pathogenic potential are more heavily distributed in high-risk regions of the world where the incidence of cervical cancer is highest [Berumen et al., 2001].

If an infection with certain variants of HPV 16 and 18 poses a greater risk for high-grade squamous intraepithelial lesion (HSIL) development, compared with an infection with an HPV 16 or -18 prototype, than this association should also be observed with the development of low-grade squamous intraepithelial lesions (LSIL), a condition that is generally believed to precede a high-grade cervical lesion. However, there is currently no indication that the risk for LSIL varies with infection by different HPV 16 or -18 variants [Villa et al., 2000; Xi et al., 2002]. If such an association could be demonstrated, testing for HPV 16 or -18 variants may improve the management of low-grade abnormal cytological results and avoid unnecessary referrals to colposcopy.

This paper presents the results from a prospective cohort study of the natural history of HPV infection and cervical neoplasia in a population of university students, in Montreal, Canada. The objectives were to evaluate the association between HPV status and LSIL and to estimate the average duration of an LSIL among women with low-risk HPV, high-risk HPV or HPV-16/18 variant infections.

METHODS

Subjects

Female students attending either the McGill or the Concordia University Health Clinic, in Montreal, Quebec, Canada, were invited to participate if they intended to remain in Montreal for the next two years and had not required treatment for cervical disease in the

last 12 months. Recruitment was initiated in November 1996 and accrual was completed in December 1998.

Procedures

All eligible women were asked to return to the clinic every 6 months over a period of 2 years, for a total of 5 visits. The participating clinics had their respective Research Ethics Board approve the study protocol. At each visit, a questionnaire was completed and endo- and ectocervical cells from the uterine cervix were collected with two Accelon cervical biosamplers (Medscand Inc., Hollywood, Fla.). A Pap smear was prepared with the first sampler and the remaining cells along with cervical cells collected with a second sampler were used for HPV DNA testing. At enrollment, information from a detailed, self-administered, questionnaire was obtained on potential risk factors for HPV infection. A follow-up questionnaire, designed to measure changes in recent sexual practices and other lifestyle factors, was completed at each subsequent visit.

HPV DNA detection

HPV DNA testing by a polymerase chain reaction in this study has been described previously [Richardson et al., 2003]. Five μl of DNA purified with the use of QUIAamp columns (QUIAGEN Inc., CA. USA) [Coutlée et al., 1999] was first amplified for β-globin DNA with PC04 and GH20 primers to demonstrate the absence of inhibitors and the integrity of processed DNA [Bauer et al., 1991; Coutlee et al., 2002]. Specimens that were β-globin-positive were further tested with the L1 consensus HPV primers MY09/MY11 and HMB01 with Amplitag Gold (TaqGold; Perkin-Elmer-Cetus, Norwalk, CT) and the line blot assay (Roche Molecular systems, CA) for the detection of 27 genital HPV genotypes [Gravitt et al., 1998; Coutlee et al., 2002]. HPV types were analyzed individually or in groups according to a classification based on their oncogenic potential. High-risk HPV (HR-HPV) types included those genotypes that are most frequently found in cervical tumours: HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68. All other individual types that were identified with the line blot assay were classified as low-risk HPV (LR-HPV) types: 6, 11, 26, 40, 42, 53, 54, 55, 57, 66, 73, 82, 83 and 84 and MM9 [Bosch et al., 1995].

Variant analysis

Clinical samples that were positive for HPV types 16 or -18 were further analyzed for HPV 16 and -18 variants by PCR-sequencing. A segment within the LCR region was amplified with primers specific for HPV 16 and -18. The latter region corresponds to a hypervariable genomic segment containing the enhancer of HPVs [Ho, 1991; Chan et al., 1992; Ho et al., 1993; Myers et al., 1995; Yamada et al., 1995]. HPV 16 and HPV 18 LCR primers flanked nucleotide positions 7478-7841 [Chan et al., 1992] and 7485-7805 [Ong et al., 1993], respectively [Chong et al., 1990] and amplification reactions were performed with a standard protocol on 5 µl of processed sample with 2.5 units of Expand High Fidelity PCR enzyme (Boehringer Mannheim, Laval, Qc.). When amplicons generated faint bands on gel electrophoresis, 10 units of *AmpliTaq Gold* DNA polymerase (Roche Molecular Systems, Mississauga, Ont.) were used instead of the Expand mixture. PCR-amplified HPV 16 or HPV 18 DNA fragments were purified with the QIAquick gel extraction kit protocol (Quiagen Inc, Mississauga, Ont.). Direct double-stranded PCR-sequencing was performed with the fluorescent cycle-sequencing method (BigDye terminator ready reaction kit, Perkin-Elmer) on an ABI Prism 3100 Genetic Analyzer system. Sequence variations that had not been described previously were confirmed by a second PCR-sequencing reaction.

If the presence of multiple variants was suspected on the electrophoregram, purified PCR products were cloned using the TOPO TA cloning kit (Invitrogen) [Mayrand et al., 2000], with the pCR2.1 TOPO vector and competent *E. coli* TOP10 strain. Ten clones containing the HPV LCR fragment were identified by digestion with restriction enzymes *Hind*III and *Xho*I. The plasmid DNA from the transformed clones was purified using the QIAprep Spin Miniprep system (Quiagen Inc.) according to the manufacturer instructions and then sequenced.

HPV variant sequences were compared, using the BLAST sequence analysis from the Genetic Computer Group (GCG) [Altschul et al., 1990] to sequences of known HPV variants or prototypes available (http//hpv-web.lanl.gov [Myers et al., 1995] and GenBank). Isolates with a DNA sequence different than that of the prototype were

classified as non-prototype variants. Sequences from unknown non-prototype variants were aligned with known variants of various lineages using the Clustal Multiple Sequence Alignment 1.8 program, to further classify variants into the appropriate lineage. The lineages were designated as E (European), As (Asian), AA (Asian-American) and Af (African). For data analysis purposes, variant branches that did not belong to the European lineage were later grouped into a Non-European group (NE).

Statistical Methods

Duration of different HPV infections was estimated with the Kaplan-Meier method. Women were classified as HPV negative if they were never positive for HPV throughout the study. The LR-HPV category included those women who only had a LR-HPV infection at any visit. Women who were positive for any HR-HPV type, other than HPV 16 or HPV 18, at any visit were included in the HR-HPV category. Women who had an HPV 16 or HPV 18 infection at one or more visits were separated into 2 groups, based on their variant status. A woman was classified as having a Non-European variant, if an NE variant was detected at one or more visits. Otherwise, she was classified with having a European variant infection.

The association between incident LSIL and different classifications of HPV infection or variants of HPV 16/18 was estimated using logistic regression. Models were adjusted for *a priori* selected potential confounders including age, because it is strongly associated with HPV acquisition [Herrero et al., 2000], and race, because variant HPV types may be distributed differentially according to ethnicity [Beckman et al., 1994; Sjalander et al., 1995]. Age was entered in the model as a continuous variable, after assessing the linearity assumptions. Race was categorized as White (reference group), Black, Hispanic and Asian.

The Kaplan-Meier method was used to estimate the cumulative probability of acquiring an incident LSIL. Similar methods were used to estimate the proportion of women who cleared their LSIL, by considering their index visit, when first diagnosed, as time zero. Women who had a prevalent or incident LSIL infection (n=43) at anytime during the study except at the last visit (visit 5) and had completed at least one subsequent visit after

infection were included in the analysis of LSIL clearance. Time to event was defined as the time between the first visit when a woman was diagnosed with an LSIL and the first subsequent visit when her cervical cytology was normal again. Those women who completed the study before clearing their LSIL or were lost (i.e. progression to HSIL or loss to follow-up) were censored at their last available visit. The four (9.3%) women who did not test positive for HPV at any visit preceding or concurrent with their LSIL diagnosis were excluded from the analysis, since it is generally believed that HPV is necessary for LSIL development [Moscicki et al., 2001]. The one-year cumulative clearance rates and median duration of the lesion were estimated with the actuarial Kaplan Meier method, separately for each HPV group. The evaluation of statistically significant differences among LSIL clearance curves for the various HPV groups was tested with the log-rank test. All hypotheses were tested at the 0.05 significance level and statistical analyses were conducted with SPSS version 11.0.

Results

There were a total of 621 women in the cohort and the average length of follow-up was 22 months (range: 0 to 48 months). A total of 424 women (68%) completed all 5 visits. Of the 621 women in the original cohort, 17 (2.7%) had an LSIL and 1 (0.2%) had an HSIL at enrollment. Thus, 603 women were eligible for the analysis of HPV status and incident LSIL. The distributions of selected baseline demographic and lifestyle factors for the cohort participants who remained free of cytological abnormalities throughout follow-up and for those women with LSIL are presented in table 8.1. Most characteristics were similarly distributed between the subjects who had normal cytology and those women who had an LSIL diagnosis during the study. However, as expected, women with an LSIL had a greater proportion of lifetime number of sexual partners and a greater occurrence of HPV infections compared to women with normal cytology.

The majority of cytological abnormalities in this cohort were LSIL's, and the HPV positivity increased with increasing cytological severity. There were 7 cases of ASCUS (atypical squamous cells of undetermined significance) in the cohort of which 86% were HPV positive. Of the 43 women with an LSIL, 91% had at least one specimen that was

HPV positive during the study. Four women had a high-grade SIL during the study and they all had a concurrent HPV infection (100%). However, so few cases of HSIL made it impossible to analyse the HSIL group separately. Thus, it was decided to exclude the HSIL cases from the analyses of SIL and focus only on the women who had an LSIL, of which one case progressed to HSIL.

The European group of variants were the most frequently detected HPV 16 or HPV 18 infections, and the distribution of specific HPV 16 or -18 variants throughout the study are presented in table 8.2a and 8.2b, respectively. The most common HPV-16 variants were G-1 and prototype infections and represented 50% and 29% of all 103 HPV-16 infections, respectively. Of the 43 visits with HPV-18 detected, the most common HPV-18 variants were the B18-2 variants (23%) and the J18-1 variants (21%). Three women with HPV 16 had different variants at the same visit and one woman had different HPV-16 variants at different visits. However, all the variants belonged to the European lineage. None of the women with HPV-18 infections experienced a co-infection with a second HPV-18 variant infection at the same or a different visit. However, of the 43 women with an HPV-18 infection, 13 women (30%) also had an HPV-16 infection at some visit. Of those, 8 women (62%) had a co-infection with HPV-16 and HPV-18 at the same visit of which 3 had an HPV-16 E variant and an HPV-18 NE variant simultaneously. The remaining 5 women had both HPV-16 and HPV-18 E variants at the same visit.

The median duration of incident HR-HPV infections (11 months), not including HPV 16 or -18, was longer than other HPV-type infections, which persisted for 9 months or less (table 8.3). When duration was estimated for both prevalent and incident HPV infections the HR-HPV types, excluding the HPV 16 or -18 variants were still the most persistent (14 months) while the Non-European variants cleared the most rapidly (7 months).

There were 26 (4.3%) women with a new diagnosis of LSIL over a two-year period, of which two were HPV negative at and before the diagnosis. The remaining 24 women with an incident LSIL had a concurrent HPV infection at the time of LSIL diagnosis. The incidence rate for acquisition of an LSIL was 2.4 per 100 woman-years (table 8.4). Figure

8.1 presents the cumulative risk of LSIL acquisition among women with specific, concurrent HPV-type infections. Women with an HPV 16/18 NE variant infection had the greatest one-year probability of acquiring an LSIL (16.7%) compared to women with a LR-HPV (6.5%), a HR-HPV (6.3%) or an HPV 16/18 E variant infection (1.0%) (figure 8.1). However, differences among the four Kaplan-Meier curves were not statistically significant (log-rank test p-value=0.48).

Women with a LR-HPV infection, a HR-HPV infection (excluding HPV 16 or -18) or an HPV-16/18 infection were all more likely to have an LSIL compared to women without an HPV infection. However, when HPV 16 and 18 variants were divided into European and Non-European groups, infection with an HPV16/18 NE variant appeared to be more strongly associated with LSIL than other HPV type infections (table 8.5) but the precision of the estimates precluded any inferences based on statistical significance.

Discussion

The progression of an HPV infection to invasive cervical cancer is thought to be a multi-step process, and a combination of molecular factors may be important in determining risk [Giannoudis & Herrington, 2001]. If some molecular variants of HPV are preferentially associated with cervical neoplasia, it may explain why some HPV infections progress to HSIL or cervical cancer while others do not.

At least four studies have recently reported results from sequencing the LCR and E6 regions of the HPV and classified variant status according to European and Non-European [Villa et al., 2000; Berumen et al., 2001; Hildesheim et al., 2001b] or prototype-like (PL) and non-prototype-like (NPL) [Xi et al., 2002] status. All four studies observed an increased risk for HSIL or cervical cancer associated with having an NE (or NPL) variant, with risks for HSIL ranging between 2.7 - 3.4, among women with an NE variant compared to women with an E variant, and relative risks of cervical cancer between 11.0 and 27.0. Villa and collaborators [Villa et al., 2000] studied the risk of SIL in a cohort of Brazilian women, according to different levels of HPV infection, and observed that women with an HPV 16 or -18 NE variant were nearly 100 times

(OR=99.7, 95%CI: 32.1, 308.9) more likely to have any SIL compared to women who were HPV negative, while women with an E variant infection were 45 times (OR=45.5, 95% CI: 18.3, 113.3) as likely to have any SIL compared to women without an HPV infection.

However, there was no indication that the risk for LSIL varied with infection by different HPV variants, according to at least two studies [Villa et al., 2000] [Xi et al., 2002]. The majority of molecular variants in our cohort belonged to the European lineage and only 10% of all the HPV 16 or HPV 18 infections were positive for Non-European branchs. Therefore, many of our analyses were limited by the small number of events, which is reflected by wide confidence intervals.

It appears that the majority of persistent, same-type infections are the original infection and not a re-infection, given that virtually all (97%) of the persistent HPV 16 or -18 infections had the same variant throughout follow-up. The incident NE variant infections were less persistent than the E variant infections in our cohort, although the estimate for the median duration was less precise than the mean duration and the confidence intervals overlapped considerably. Despite appearing less persistent, there was the suggestion of a stronger association, based on the odds ratio estimate, between NE variants and LSIL than between E-variant or other HPV-type infections and such lesions. This finding suggests that some HPV 16/18 NE variants are more pathogenic then other HPV types or variants because they are associated with a shorter period of duration but with a higher probability of LSIL occurrence.

However, the challenge of studying the association between HPV variants and SIL development is based on the inherent difficulty in evaluating the individual effects of different variants. Historically, researchers have had to group HPV types together based on presumed oncogenic risk [Bosch et al., 1995], or phylogenetic likeness inferred by DNA sequence homology [Thomas et al., 2000] due to the rare distribution of many HPV genotypes in different populations. Similarly HPV-16 and HPV-18 variants are generally grouped together based on geographical co-evolutionary lineages [Ho, 1991], in order to gain statistical power. In our cohort, 18 different HPV-16 variants and 11 different HPV-

18 variants were identified. The very small number of women with cervical lesions in our cohort necessitated the combining of variants into a European and a Non-European group and the European category contained the majority (83%) of the individual variants. Thus, despite the apparent increased risk associated with NE variant infections, adjusted for age and race, we cannot be certain that the effect is not just an artifact of the classification of variants that effectively led to a pooling of effects of individual variants. Moreover, the parameter estimates were very unstable, reflected in the exceedingly wide confidence intervals, making it difficult to reach any firm conclusions.

It could be argued that LSILs are not an ideal endpoint for studying the effects of HPV variants on neoplastic progression since low-grade lesions are often synonymous with an HPV infection. However, different viral variants may have altered biological properties that can influence the efficiency of transcriptional regulation and modify viral transforming abilities thereby affecting the ability to persist [Hildesheim et al., 2001b; Xi et al., 2002]. Furthermore, it is generally agreed that cervical neoplasia is a gradient of disease that progresses from low-grade severity to high-grade severity and the maintenance of a persistent HPV infection is required for establishment of a malignant lesion [Nobbenhuis et al., 2001]. Therefore, if certain HPV 16 and 18 variants are more pathogenic then other HPV types or variants, then specific stages in the natural history of HPV, including preneoplastic progression to both LSIL and HSIL, should be studied, so as to clarify the role that viral variants may play in the transition from a persistent HPV infection to different stages of cervical intraepithelial neoplasia.

There are undoubtedly other factors, apart from HPV type and persistence that will influence the regression or progression of an LSIL, including the continuous expression of E6 and E7 viral genes in cycling cells, integration of viral DNA into the host cell chromosome and host factors such as HLA genotype and polymorphism of cell cycle regulating genes [Giannoudis & Herrington, 2001]. Nonetheless, the clinical relevance of a persistent HPV16 or HPV 18 infection with NE variants should be further investigated. One large scale randomised clinical trial recently concluded that HPV testing for the management of LSILs is futile, since most LSILs are positive for HR-HPV types [The

ALTS Group, 2000] and therefore, carries a low positive predictive value. However, if future studies can help identify the persistent variants associated with progression from the variants associated with regression of cervical lesions the subject of LSIL management with HPV testing may warrant a re-visit.

Acknowledgments

This study was funded by a grant from the Canadian Institutes of Health Research (CIHR) (MT-13649). HR is recipient of a Doctoral Scholarship, MA of a Health Scientist Award, and EF of a Distinguished Scientist Award, all from the CIHR. FC is recipient of a Chercheur Boursier National from the FRSQ. The authors are indebted to Suzanne Dumais and the staff at the McGill and Concordia Student Health Services who helped with subject accrual and follow-up. The authors also thank Roche Molecular Systems who generously supplied us with all the reagents necessary for HPV genotyping.

Table 8.1
Distribution of selected variables at baseline among participants with normal cytology and low-grade SIL during the study

Selected risk factors	Normal cytology at all visits N=568 (%)	Cumulative LSIL N=43 (%)
Age		
17-20 years	172 (30.3)	12 (27.9)
21-23 years	208 (36.6)	17 (39.5)
24-26 years	91 (16.0)	9 (20.9)
27+ years	97 (17.1)	5 (11.6)
Race		
White	469 (82.6)	34 (79.0)
Black	26 (4.6)	3 (7.0)
Hispanic	16 (2.8)	1 (2.3)
Asian	57 (10.0)	5 (11.6)
Smoking status		
Never	338 (59.5)	28 (65.1)
Former	94 (16.5)	4 (9.3)
Current	136 (23.9)	11 (25.6)
Age at first sexual intercourse		
19+ years	159 (28.0)	11 (25.6)
16-18 years	283 (49.8)	25 (58.1)
<16 years	126 (21.2)	7 (16.3)
Number lifetime sex partners		
1 lifetime partners	139 (24.5)	7 (16.3)
2-4 lifetime partners	181 (31.9)	11 (25.6)
5+ lifetime partners	248 (43.7)	25 (58.2)
Oral contraceptive use		
Never in adult life	132 (23.2)	13 (30.2)
Sometimes in adult life	60 (10.6)	4 (9.3)
Always in adult life	376 (66.2)	26 (60.5)
***HPV status:**		
Negative	289 (50.9)	4 (9.3)
Positive	279 (49.1)	39 (90.7)

** Cumulative HPV positivity throughout study*

Table 8.2a: Distribution of women with specific HPV-16 variants at one or more visits

No. women with each variant	7478	7479	7483	7487	7489	7490	7494	7502	7505	7519	7533	7550	7557	7562	7565	7566	7576	7595	7622	7623	7628	7634	7637	7641	7645	7646	7659	7665	7667	7687	7711	7712	7727	7728	7741	7747	7753	7762	7771	7776	7779	7784	7790	7797	7800	7816	7824	7832	7835	7837	7840	Variants	HPV-16 lineage
	A	C	A	G	T	T	T	T	A	G	C	G	A	C	G	T	C	T	T	G	T	A	A	T	C	A	A	A	C	C	T	T	A	A	T	A	T	C	A	A	T	C	C	G	C	C	G	G	A	A	G	Reference nucleotide	
	T	A	C	A	C	A	C	C	G	A	A	A	G	T	A	G	T	C	C	A	C	C	G	G	T	G	C	G	T	A	G	A	T/C	C	G	G	C	T	G	G	C	T	T	C	A	T	A	T	T	G	A	Mutation	
30 (19)*																																																				Prototype	European
52 (27)																																																				G-1	European
2 (1)																																																				SA-9	European
3 (2)																																																				G-10	European
1 (1)																																																				PF16-1	European
1 (1)																																																				PF16-2	European
1 (1)																																																				PF16-3	European
1 (1)																																																				PF16-4	European
1 (0)			T																																																	PF16-5	European
1 (1)																																																					PF16-6	European
1 (1)																																																					PF16-7	European
1 (1)																																																					PF16-8	European
1 (1)																																																					PF16-9	European
1 (0)																																																					PF16-10	European
3 (1)									C																																											U33068	European
1 (1)																																	C																			IND-8	Asian/American
1 (1)																																																					U34089	African-2
1 (1)																																	C																			PF16-11	African-1 & Asian/American

*Note: Total of 103 women with HPV-16. * Value in bracket () represents the proportion of women with the same variant detected at two or more visits. Note: Four women with HPV 16 had different variants at the same or a different visit.*
Participant A: visit 1=G1; visit 3=G1 & PF16-2; visit 4=PF16-2; Participant B: visit 4=PF16-3 & G1; visit 5=PF16-3;
Participant C: visit 1=G1; visit 2 & 3=prototype; Participant D: visit 1=PF16-7; visit 2=PF16-7 & prototype; visit 3=prototype

Table 8.2b: Distribution of women with specific HPV-18 variants at one or more visits

No. of women with each variant	7486	7496	7507	7512	7527	7528	7529	7530	7536	7549	7551	7563	7564	7567	7592	7643	7651	7658	7670	7681	7698	7704	7717	7726	7730	7764	7783	Variants	HPV-18 lineage
	C	C	A	G	A	A	C	T	C	C	C	G	A	A	T	T	T	A	A	T	T	T	A	C	C	C	A	Reference nucleotide	
	T	G	G	A	C	C	A	C	T	T	T	A	C	C	C	G	C	C	T	C	G	C	C	T	A	T	C	Mutation	
0																												Prototype	European
8 (6)*																												G18-1	European
10 (4)																												B18-2	European
1(0)																												H18-4	European
3 (1)																							G					Sc18-2	European
1 (0)																												Sc18-4a	European
3 (2)																												T18-9	European
1 (0)																												HV18-1	European
3 (1)																												HV18-2	European
2 (0)																												HV18-3	European
9 (5)																												J18-1	Asian/American
2 (1)																												NY18-1	Asian/American

*Note: Total of 43 women with HPV-18. *Value in bracket () represents the proportion of women with the same variant detected at two or more visits. None of the women with HPV-18 experienced a co-infection with a second HPV 18 variant at the same or a different visit.*

Table 8.3

Average duration of specific HPV-type or HPV-16/18 variant infections

HPV group or variant	Duration of incident infections (months)		Duration of prevalent & incident infections (months)	
	Mean (95%CI)	Median (95%CI)	Mean (95%CI)	Median (95%CI)
Low-risk	11.4 (9.3, 13.5)	8.6 (4.0, 13.3)	12.8 (10.6, 14.9)	11.8 (8.1, 15.5)
High-risk*	12.9 (11.1, 14.7)	11.4 (8.2, 14.7)	15.6 (13.7, 17.6)	13.8 (10.7, 16.9)
HPV 16/18 E-variants	11.3 (8.6, 14.0)	7.9 (3.0, 12.7)	11.9 (10.2, 13.6)	10.8 (6.7, 14.9)
HPV 16/18 NE-variants	7.0 (5.2, 8.8)	7.0 (6.6, 7.3)	9.6 (6.7, 12.4)	7.0 (6.9, 7.1)

**Excluding HPV 16 or -18 infections*

Table 8.4
Incidence rates for LSIL according to cumulative exposure to HPV infection throughout follow-up

Cumulative HPV Status	No. women at risk*	# events / # person-years	LSIL: IR per 100 person-years (95% CI)
HPV negative	261	2/502.5	0.4 (0.2, 1.0)
LR-HPV	62	5/120.4	4.2 (0.6, 7.8)
HR-HPV**	114	10/222.7	4.4 (1.6, 7.2)
HPV 16/18 E-variants	111	7/222.1	3.2 (0.9, 5.5)
HPV 16/18 NE-variants	12	2/21	9.5 (0.0, 22.7)
All HPV positive women	299	24/586.2	4.1 (2.5, 5.7)
All women in study	560	26/1093.2	2.4 (1.5, 3.3)

** Number of women at risk included those women with normal cytology at enrollment, and with ≥ 1 return visit.*

*** Excluding HPV 16 or -18 infections*

Table 8.5
Association between HPV infection and incident LSIL: crude and adjusted for age and race

HPV Status (LSIL%)	Crude OR (95% CI)	Adjusted[5] OR (95% CI)
Negative 293 (0.7)	REF	REF
LR-HPV[1] 64 (7.8)	12.3 (2.3, 64.7)	11.6 (2.2, 61.8)
HR-HPV[2] 120 (8.3)	13.2 (2.8, 60.9)	13.5 (2.9, 62.9)
HPV16/18 E-variant[3] 114 (6.1)	9.5 (1.9, 46.2)	10.6 (2.1, 52.4)
HPV16/18 NE-variant[4] 12 (16.7)	28.9 (3.7, 226.6)	38.1 (4.7, 310.9)

[1] *Exclusively LR-HPV infections.*
[2] *HR-HPV infections excluding HPV 16 or -18.*
[3] *European variants*
[4] *Non-European variants for HPV 16 or -18.*
[5] *The OR's are adjusted for age and race in a multiple logistic regression model.*

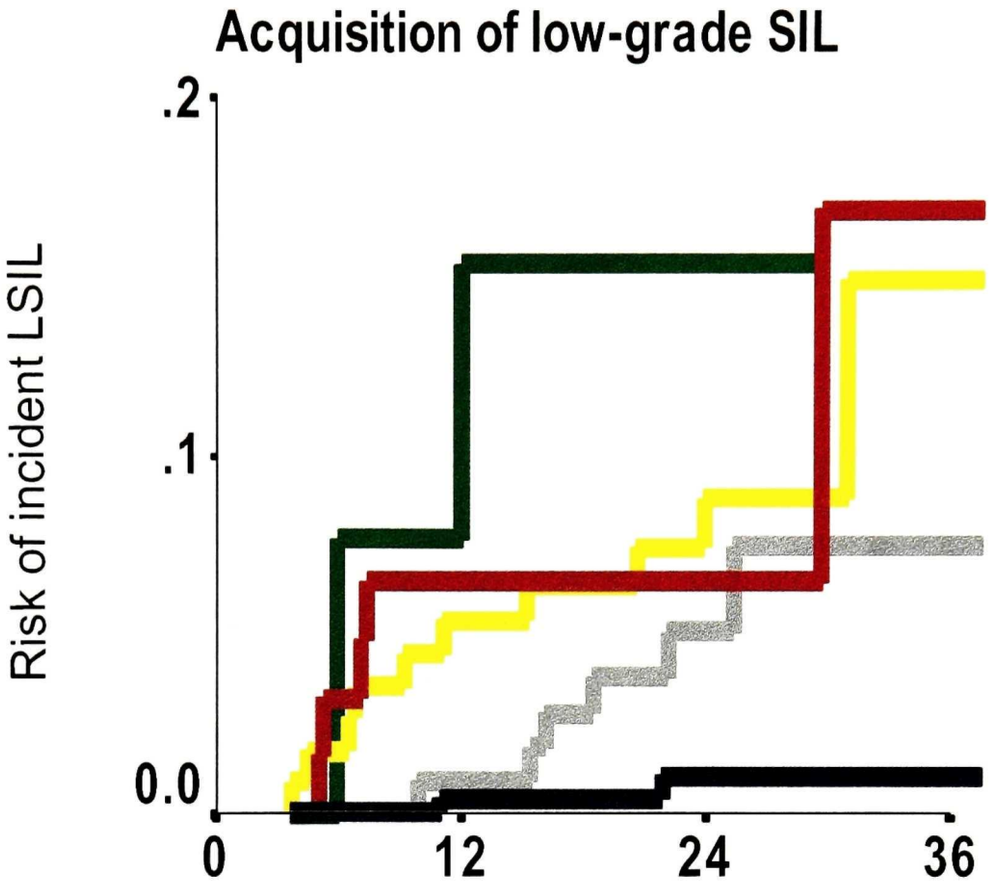

Figure 8.1

Legend for figure 8.1

- HPV negative throughout study
- LR-HPV only during f-up
- HR-HPV (excl. HPV 16/18) during f-up
- European HPV-16/18 variant during f-up
- Non-European HPV 16/18 variant during f-up

Legend for Figure 8.1:

Cumulative probability of low-grade SIL among women who were HPV negative during the study or had a specific HPV-group infection. The separate curves included the following subset of subjects:

a) Subjects who were HPV negative during the study (n=261)
b) Subjects with exclusive LR-HPV infections (n=62)
c) Subjects with HR-HPV infections, excluding HPV 16/18 (n=114)
d) Subjects with HPV 16/18 European variants (n=111)
e) Subjects with HPV 16/18 Non-European variants (n=12)

8.2 Additional analyses

The severity of an LSIL was investigated according to the most oncogenic HPV type that was present concurrently at the same time that an LSIL was first diagnosed. For the purposes of this analysis, infections with an HPV 16- or- 18 Non-European variant were assumed to be the most severe HPV infections, while an infection with a LR-HPV infection was the least severe. Time-to-regression of an LSIL (cytology diagnosis of 'normal'), stratified by type of HPV infection, was estimated with the Kaplan Meier method. The average duration of LSILs with different HPV infections is presented in table IV.i in Appendix IV. The shortest persisting LSIL was associated with an HPV 16 or -18 NE variant and the longest persisting LSIL was associated with an HPV 16 or -18 E variant. However, time-to-LSIL regression was only based on two observations in the NE group. Therefore, little can be concluded from this analysis.

CHAPTER 9: DISCUSSION

This chapter is based on a global summary of the main results from the manuscripts included in the thesis and an evaluation of the impact of the potential methodological limitations of the study design. The study rationale accompanied with a brief summary of the epidemiology of cervical HPV infections are provided to highlight the gaps in the literature that this project was designed to address.

9.1 Rationale

Certain types of oncogenic HPVs are considered a necessary cause of cervical cancer [Walboomers et al., 1999]. Alarmingly, cervical HPV infections are also the most common sexually transmitted infections in most of the world. However, epidemiological studies, to date, have shown that only a small fraction of women infected with oncogenic HPV types will eventually progress to HSIL and cervical cancer, and the remaining cases of subclinical HPV infection are transient and disappear naturally [Moscicki et al., 1998; Franco et al., 1999; Schlecht et al., 2001].

Given that most HPV infections appear to resolve spontaneously, HPV infections are not considered a sufficient cause of cervical cancer. It has been assumed that other factors, in conjunction with HPV, modulate the risk of transition from cervical HPV infection to cervical malignancy [Castellsague et al., 2002b]. Little is known, however, about the occurrence and determinants or environmental co-factors of persistent HPV infections. Considering that there is now an ongoing debate concerning whether HPV testing should be added to existing cervical cancer screening programs, it is important that the natural history of HPV infections be well understood and the issues related to viral persistence be addressed by epidemiological studies.

9.2 Prior Knowledge

Molecular epidemiology studies using PCR to detect viral DNA have observed that 7-50% of sexually active women are positive for HPV, and HR-HPV types are thought to be more prevalent than LR-HPV types [Kjaer et al., 1993; Bosch et al., 2002]. Prior to the

start of this study, in 1996, there was no published data on HPV incidence or persistence. Since then, a few studies have reported preliminary estimates of HPV incidence from data collected prospectively. HPV incidence rates appear to range from 16 to 47 new cases per 100 woman-years [Van Doornum et al., 1994; Franco et al., 1999; Thomas et al., 2000; Giuliano et al., 2002a] and the estimated 2 year cumulative incidence varies between 18% and 39% [Ho et al., 1998; Franco et al., 1999; Moscicki et al., 2001; Winer et al., 2003].

Among studies in which more than 2 visits were considered in the determination of HPV status, the proportion of HPV persistent infections has been found to range between 11% in middle aged women [Franco et al., 1999] to just over 50% among women in their early to mid twenties [Ho et al., 1995; Giuliano et al., 2002a]. The median duration of newly acquired or prevalent HPV infections appears to range from 8 months to 14 months in young and middle aged women [Ho et al., 1998; Franco et al., 1999; Woodman et al., 2001; Giuliano et al., 2002a] and HR-HPV infections appear to persist twice as long as LR-HPV infections. A co-infection with other HPV types at the same visit is extremely common, particularly in young women, and has been observed to represent 15% to 40% of all HPV infections [Rousseau et al., 2000; Woodman et al., 2001]. However, at least one group that studied persistence of HPV 16 variants did not observe co-infections with different HPV 16 variants at the same or a different visit [Villa et al., 2000].

The risk of HPV infection is strongly associated with age and sexual activity, although a number of recent studies have observed that markers of lifetime sexual activity may be more strongly associated with HR-HPV infections [Franco et al., 1995; Kjaer et al., 1997; Rousseau et al., 2000; Richardson et al., 2000; Chan et al., 2002]. Whether similar associations exist between sexual activity and incident HR- and LR-HPV infections remains to be confirmed. Furthermore, it is still not certain that every woman has an equal risk of acquiring an HPV infection, even if she is sexually exposed to the virus. Other variables such as oral contraceptive use, parity, tobacco consumption and a history of STD's have been sporadically identified as co-factors that may help facilitate HPV acquisition or successful transmission to the target tissue. The inconsistencies in the

literature, to date, may be due to the absence of data on recent exposure to putative co-factors or may also be due to the possibility that the effect of some of these co-factors act differentially according to the oncogenic characteristics of the virus.

The strongest predictors of a persistent HPV infection, thus far, include HPV type, viral load and the host immune response [Maciag et al., 2000]. However, very few studies have observed an association between environmental co-factors and persistent HPV infection. Similarly, while persistent HPV infection is an important predictor of SIL, whether viral variants (i.e., intra-type variability) contribute to the risk of transition from a persistent HPV infection to cervical neoplasia is still not clear.

9.3 Summary of results

The following section is a brief summary highlighting the results from the four manuscripts included in the thesis. When appropriate, specific issues or explanations that had to be left out of the manuscripts for the sake of brevity are also discussed.

Manuscript I

The first manuscript in this thesis aimed to describe the natural history of HPV infections, independent of environmental co-factors. The two-year cumulative incidence for any HPV infection in our cohort was 36%, closely resembling the incidence of HPV in similar cohorts of young women in North America and Britain. The incidence rate for any HPV infection was 22.8 per 100 woman-years (95% CI: 21.0, 24.6). The incidence rates for HR- (IR=16.8, 95% CI: 15.3, 18.3) and LR-HPV infections (IR=14.9, 95% CI: 13.6, 16.2) did not appear to differ substantially, unlike the prevalence proportions which were higher for HR-(21.8%) than LR-HPV infections (14.8%). The three most frequently acquired type-specific infections were HPV 16 (12%), HPV 51 and HPV 84 (8%).

Previous studies that generally measured duration of both prevalent and incident HPV infections observed that LR-HPV types clear faster than HR-HPV types. However, studying prevalent infections makes it difficult to estimate the real average duration of infection, since there is no way of telling how long the prevalent infections had persisted

before detection at enrollment. In contrast, we only estimated the duration of incident infections, and observed that there was very little difference in the average duration of HR- and LR-HPV infections. In our cohort, the median duration of the most type-specific persistent HR-HPV infection (13.2 months) was similar to the median duration of a LR-HPV infection (12.3 months).

The discrepancy in results from our study and some of the other cohort studies that have estimated average duration of HPV infections may be explained, in part, by different age distributions of the cohorts and different detection methods used. For example, the San Francisco cohort included very young women (13-20 years old), and was constrained by less sensitive detection methods that were available in the early 1990's [Moscicki et al., 1998]. Duration varied significantly depending on how clearance was defined. It was only with the most rigid model, that required women to have 3 consecutive HPV negative visits to be classified as 'cleared', (in an attempt to reduce misclassification of the outcome), did the estimated median duration of HR-HPV (~ 12 months) and LR-HPV (~10 months) infections approximate our results [Moscicki et al., 1998]. Two other cohorts, one in Brazil, amongst women aged 18-65 [Franco et al., 1999] and the second, amongst women in Arizona aged 18-35 [Giuliano et al., 2001], observed the median duration to be less then 10 months for a HR-HPV infection and to be closer to 5 months for LR-HPV infections. The women were older on average in these two cohorts then the women in our study and it is possible that the cohort members may already have acquired varying levels of immunity against HPV, from past exposure to different HPV types.

The cumulative proportion of women with co-infections, defined as a LR and a HR-HPV type detected at the same visit, was extremely high (38%). The overall proportion of co-infections with any two HPV types at the same visit was substantially higher (52.3%).

Manuscripts II and III

The objective of manuscript II was to identify co-factors for acquisition of LR- and HR-cervical HPV infections, independent of sexual activity. Manuscript III focused on the identification of putative determinants that could significantly contribute to the rate of

clearance of mutually exclusive LR- and HR- cervical HPV infections. Because of the many overlapping risk factors that were explored in both manuscripts, the results from these two chapters will be summarized together.

There are a number of conceivable mechanisms that may either facilitate or hinder HPV acquisition and persistence. Factors that may cause cervical irritation and affect the integrity of the cervical squamous epithelia such as the use of tampons or a current *Chlamydia* or HSV-2 infection could play a facilitating role in the transmission of HPV to the target basal epithelial cells in the cervix. Alternatively, other factors may hinder HPV transmission including precautionary practices such as the use of vaginal lubrication (to minimize vaginal abrasion), condoms, and washing after sexual activity. Another possible mechanism for acquisition or persistence may be through the attenuation of the immune response to the virus, thus facilitating propagation and persistence of a cervical HPV infection. Factors that might have such endogenous or hormonal influences on the host immune system include tobacco metabolites, alcohol, oral contraceptives and certain dietary nutrients. This was the rationale for exploring the relationship between HPV and the following factors described below.

Age and sexual activity

As expected, age and sexual activity were associated with HPV acquisition, irrespective of type. However, age at first sexual intercourse was not associated with acquisition of either type of HPV infection. It has been suggested that age at first intercourse may be a proxy for first exposure to HPV [Schiffman & Brinton, 1995; Deacon et al., 2000] and would therefore be a better predictor of a latent, persistent HPV infection rather than a new infection that is likely transient. However, despite positive associations observed between HPV persistence and younger age at first intercourse in some studies [Ho et al., 1998; Kruger-Kjaer et al., 2001; Moscicki et al., 2001], we were not able to corroborate these results. Furthermore, we failed to observe a significant association between age and HPV clearance. Nonetheless, it should be emphasized that there was limited variation in age in this cohort and the vast majority (>90%) of women were younger than 30 years of age.

Race

Race was not related to HPV acquisition, but was associated with clearance of a HR-HPV infection. Women from African (Black) descent were significantly more likely to clear their HR-HPV infection than Caucasian women (HR=2.86; 95% CI: 1.18, 6.95). These results are not in agreement with a cohort study of university women in New York where the authors observed that women of African-American or Hispanic descent were at greater risk for a persistent HPV infection compared to women who were predominantly Caucasian [Ho et al., 1998]. The interest in race is driven by recent research that has reported on marked ethnic variations in the distribution of P53 polymorphisms [Beckman et al., 1994; Sjalander et al., 1995] in the host genome and the frequency with which certain polymorphisms are associated with cervical cancer [Storey et al., 1998; Makni et al., 2000]. However, the definition of race or ethnicity is extremely challenging and, in a heterogeneous population like Montreal, is very prone to misclassification once categories are created. Furthermore, in the context of this study, the vast majority of women were Caucasian and race is more likely a proxy for other unmeasured socio-cultural factors such as attitudes towards access to healthcare and screening, and other lifestyle choices including diet, rather than differences in genetic susceptibility.

Oral contraceptives

Increased duration of OC use was significantly protective against a new LR-HPV infection and was associated with faster LR-HPV clearance (although not statistically significant). It is possible that long-term use of OCs is a proxy for long-term monogamous relationships, which could reflect less risky sexual behaviour on the part of the participant or her male partner. However, why this effect would be differential according to HPV oncogenecity is not clear, particularly since women who were current OC users were significantly more likely to acquire a HR-HPV infection then women who were not current users (table 6.2). Duration of OC use is a composite of past and current exposure. Women exposed to OCs for 5 or more years are likely to be older and more sexually experienced then women currently using OCs for one year or less. It is possible that this group of long-term OC users have already been exposed to past HPV infections

but their immune response is more efficient at recognizing and clearing LR-HPV types then HR-HPV types. It is also possible, however, that the association between recent (current) OC use and HR-HPV acquisition is explained by residual confounding due to unmeasured markers of sexual activity.

A few recent experimental studies have observed that steroid hormones including glucocorticoids and progesterone can interact with hormone-response elements in the viral long control region of HPV 16, enhancing HPV transcription and resulting in transformation of cervical cells [Mittal et al., 1993; Pater et al., 1994]. Susceptibility to different HPV type infections may be modulated by oral contraceptives that contain different concentrations of estrogen and/or progesterone. However, one of the limitations in our questionnaire was the lack of detailed information about the brand names of oral contraceptives that women used. Consequently, we were not able to evaluate the individual risk of persistence associated with specific hormone combinations of the pill.

Condoms

Our study was not able to help identify a protective effect of condom use against HPV transmission. Despite the suggestive protective effect of recent condom use against acquisition of either LR- or HR-HPV infections, the point estimates were not statistically significant and there did not appear to be a dose-response with increasing regularity of condom use. Measuring the effect of condom use is extremely challenging, since condoms can serve two functions as a contraceptive barrier and/or an STD barrier. Because we did not make the distinction, in our questionnaire, between these two functions, we could not distinguish between women who used condoms throughout each entire sexual encounter, and those women who delayed the use of condoms towards the end of a sexual encounter, to prevent becoming pregnant. Furthermore, because we only measured frequency of condom use according to three levels (never, occasionally and always), a participant might have responded that she "always" used condoms with her partner even if they were only used 90% of the time. Consequently, exposure assessment of condom use may have been misclassified.

Paradoxically, recent regular condom use was protective against LR-HPV persistence in our study. Condom use and other barrier contraceptives have not been shown to be associated with HPV persistence in previous studies [Brisson et al., 1996; Moscicki et al., 1998]. Nonetheless, regular use of condoms has been associated with a reduction in risk for high-grade squamous intraepithelial lesions among HPV positive women in two separate cohorts [Ho et al., 1998; Hildesheim et al., 2001a]. It is possible that some HPV types are more mucosotropic than others and less likely to reside on dryer surfaces of the genital epithelium. Condoms might protect against mucosotropic HPVs more effectively than against cutaneo/mucosotropic HPVs that would be found on the lower male genitalia. Therefore, while condoms may not provide 100% protection against HPV transmission, it is possible that they minimize exposure to certain HPV types and consequently, reduce the probability of concurrent or sequential co-infection, which in turn might reduce the risk of HPV persistence [Ho et al., 1998].

STDs

In the present study, after controlling for the potential confounding effect of sexual activity and condom use, the association between a recent diagnosis of *Chlamydia trachomatis* (self-reported) and acquisition of either HPV-group infection was significantly elevated. However, while it has been suggested that *Chlamydia* infections may act to irritate the cervical epithelium, thereby facilitating HPV transmission, it is very difficult to be certain that the true association was not masked by residual confounding. It is possible that a recent *Chlamydia* infection explains additional components of sexual behaviour not measured by number of sexual partners or frequency of sex. Therefore, even with adjustment for traditional markers of sexual activity some dimensions of sexual behaviour such as the level of risk behaviour of male partners may not have been captured.

An HPV infection with other sexually transmitted diseases such as HSV-2 or *Chlamydia* has been inconsistently associated with cervical carcinogenesis and women infected with both HIV and HPV are at a much higher risk of SILs than women infected with either of the viruses separately [Castellsague et al., 2002a]. Various mechanisms for the effect

modification of STDs have been proposed, including host immunosuppression or induction of an inflammatory response that could modulate the effect of HPV. However, we did not observe an association between STDs and HPV persistence (a presumed early stage of cervical carcinogenesis). Nonetheless, our measure of a recent STD exposure was based on self- reporting and prone to misclassification. There may have been women who were unaware of a past or current infection, especially with STD's such as *Chlamydia,* that are associated with silent symptoms. Obviously, laboratory detection of other STDs in biological samples would have been a more reliable measurement of recent exposure. However, at least one cohort study that tested for HPV, *Neisseria gonorrhoeae*, *Chlamydia trachomatis* and *Trichomonas vaginalis* did not observe an association with these STD's and HPV clearance [Moscicki et al., 1998].

Moscicki and collaborators [Moscicki et al., 1998] did observe that young women were half as likely to clear an HPV infection with an incident infection with vulvar condylomas [Moscicki et al., 1998]. In contrast we observed that women with a recent history of genital warts cleared their LR-HPV infections more quickly than women without genital warts. While seemingly paradoxical, our results might be explained by type-specific immunity. The majority of external genital warts are caused by HPV-6 or HPV-11, both of which are low-risk types. Humoral immunity against HPV is primarily type-specific [Roden et al., 1996; White et al., 1998], although cross-protection against other related types has been observed [Roden et al., 2000]. Thus, it is possible that only women with LR-HPV infections would benefit from a heightened immune response to HPV-6 or -11 induced genital warts (condylomas).

Tobacco

We did not observe an association between tobacco exposure, when measured in pack-years smoked, (average number of packs of cigarettes smoked per day in life), and acquisition of a new HPV infection. Similarly, two other cohort studies, one amongst young women from a university setting [Ho et al., 1998] and the other amongst women attending Family-Planning clinics [Moscicki et al., 2001], did not observe an association with tobacco use and incident HPV infections . However, results based on data that

included prevalent cases of HPV have shown smoking to either be a risk factor for overall HPV infections [Bauer et al., 1993], LR-HPV infections [Rousseau et al., 2000] or HR-HPV infections [Chan et al., 2002]. The elevated risk of HPV infection among smokers observed in some of these studies may be more correlated to persistent HPV infections rather than new infections, since prevalent as well as incident infections were included in the analyses. Our study showed that women who had smoked one to two packs of cigarettes per day for at least one year were nearly half as likely to clear their HPV infections then women who never smoked, although the association was marginally non-significant. Nonetheless, we were not able to observe a dose-response, in part perhaps, because of the small proportion of long-term, heavy smokers in this cohort.

Alcohol

Our results were in concordance with one other study that observed increased alcohol consumption to be associated with acquisition of cervical HPV infections in university students [Ho et al., 1998]. It has been suggested that alcohol may modulate the circulating levels of estrogen and the vaginal epithelium may be more estrogen responsive, including the cervix [Hankinson et al., 1995; Reichman et al., 1993]. However, despite appealing biological mechanisms that may explain a causal relationship between alcohol and HPV infection, alcohol needs to be further studied, in light of potential residual confounding by sexual activity and unprotected sex, before any firm conclusions can be made about alcohol consumption and risk of HPV infection.

The results in the present study showing that increased alcohol intake increased the rate of HPV clearance did not support previous observations that showed increased alcohol consumption was either associated with increased HPV persistence [Ho et al., 1998], or not associated with HPV clearance at all [Moscicki et al., 2001]. However, the effect of heavy drinking on HPV clearance could not be fully assessed in this study because approximately 90% of the participants drank less than 7 drinks per week. Therefore, the relationship between alcohol and increased clearance may have been confounded by other potentially protective characteristics associated with moderate alcohol intake that

were not measured in this study, such as lower levels of stress or increased involvement in team sports and physical activity.

Diet

Weekly consumption of dairy products (at least one cup of milk or one serving of cheese per week) was significantly protective (HR=0.17, 95% CI: 0.08, 0.35) against the acquisition of a HR-HPV infection. However, the absence of any dairy products from the diet is very rare and women who never consume dairy products may differ from the remaining cohort with respect to other nutritional factors because of unusual food allergies or lactose intolerance. It is possible that low levels of calcium, or other important micronutrients from dairy such as vitamin A and D, may affect the biochemical integrity of the cervical epithelium that in turn facilitates HPV incorporation into the squamous cell epithelium. However, without additional, detailed dietary information on the subgroup of women who never consume dairy products, it is difficult to speculate on the association with HPV acquisition and reduced dairy intake.

In the present study, women who ate vegetables daily cleared their HPV infections twice as quickly as women who did not. We only measured the average frequency of certain vegetables throughout the participant's adult life in the questionnaire at baseline. Food frequency questionnaires have not been shown to be very reliable, particularly if only administered once [Byers et al., 1987; Zielinski et al., 2001]. Furthermore, our questionnaire was a very crude dietary measurement instrument and was designed to capture very general eating habits, due to the exploratory nature of the study. Nonetheless, our results are in concordance with a recent cohort study of young women from Arizona that measured levels of micronutrient intake in the blood and also evaluated the association between diet and HPV persistence using the validated "Health Habits and History Questionnaire" from the National Cancer Institute [Sedjo et al., 2002b]. Results from this study showed that women with the highest level of vegetable intake in their diet were half as likely to have a persistent HR-HPV infection compared to women in the lowest quartile of exposure and that circulating levels of B_{12} in the blood was significantly protective against a persistent HPV infection.

Post-sexual ablutionary practices and hygiene products

In this study, women who usually washed within an hour after sexual intercourse were half as likely to acquire a HR-HPV infection then women who did not. Cleansing habits, including douching and washing after sex, have been proposed as explanatory variables for observed differences in HPV prevalence, but reported associations have been inconsistent [Ho et al., 1998; Richardson et al., 2000; Rousseau et al., 2000]. Similar findings were observed in this same population in an earlier study of prevalent HPV infections [Richardson et al., 2000]. Douching was a very rare practice in this cohort and the association with HPV detection could not be adequately assessed.

Our results showed that women who used tampons instead of sanitary napkins during menstruation were significantly less likely to clear their HR-HPV infection. Only one other study to date has observed a similar elevated risk for cumulative HR-HPV infection among women who used tampons instead of sanitary napkins [Rousseau et al., 2000]. While the mechanism is currently unknown, it is possible that tampons may serve to spread the viral infection to different loci, thereby increasing the area and possibly the severity of the HPV infection.

Manuscript IV

Manuscript four explored the relationship between viral variants and LSILs. The majority of persistent, same-type HPV infections were observed to be the original infection and not a re-infection; given that virtually all (97%) of the persistent HPV 16/18 variant infections were the same variant throughout follow-up. The incident Non-European (NE) variant infections were less persistent than the European (E) variant infections in our cohort. Nonetheless, there was the suggestion of a stronger association, based on the point estimate, between NE variants and LSIL (OR=38.1 95% CI: 4.7, 310.9) then between E-variants (OR=10.6 95% CI: 2.1, 52.4) and LSIL.

The overall incidence rate for LSIL among women in the cohort was 2.4 per 100 person-years, and was 4.1 per 100 person-years among women with HPV. Duration of LSILs

was estimated based on the HPV variant or type detected at diagnosis. An LSIL with a European variant appeared to take the longest to resolve with a mean duration of 9 months (5.7, 12.2). However, none of the estimates of duration, according to HPV infection, were very precise and the confidence intervals for the different HPV groups overlapped considerably. Therefore, it was not possible to identify predictors of LSIL regression with any certainty.

9.4 Methodological issues in the present study

9.4.1 Sample size

One of the limitations of the present study was the small number of exclusively LR- or HR-HPV infections in the cohort, due to the large proportion of co-infections among HPV positive participants. In order to appropriately investigate predictors of HR and LR-HPV clearance, women with an HPV infection were assigned to either the HR group or the LR group, based on the type with the longest duration. This strategy resulted in a relatively small number of women included in the LR-HPV group, since the majority of co-infections involved longer persisting HR-HPV infections. Consequently, the small sample size for women with a LR-HPV infection in the 3rd manuscript prevented the estimation of more precise parameters.

Another limitation of the study was the over-estimation of expected abnormal cytological developments in a cohort with 600 women followed over two years. Based on results from the previous pilot study, there were only half as many diagnosed cytological abnormalities as expected. Therefore, it was very difficult to evaluate the association between HPV 16 or -18 variants and incident LSIL with any precision, because of the very small number of new cases of LSIL observed in our cohort. This set of analyses was further hindered by the very low prevalence of Non-European HPV 16 or -18 variants in the cohort.

9.4.2 Loss-to-follow-up

Approximately 90% of the cohort returned for at least 2 return visits. Nonetheless, approximately one-third of the cohort was eventually lost to follow-up before the last

scheduled visit. If women with a certain exposure profile were less likely to return for a subsequent visit if they were predominantly HPV positive (or persistent), or conversely, predominantly HPV negative (or transient), then certain associations that were measured in this study would not be valid, and we would not able to predict the direction in which the estimated hazard ratio or relative risk would be biased.

However, none of the participants in our study were informed about their HPV status until the completion of the study, so it is unlikely that loss to follow-up was influenced by the primary study outcome. Moreover, the distribution of exposure variables measured at baseline for those women lost to follow-up did not differ substantially from the distribution of exposure variables measured at baseline for the subset of women who completed all five visits. Therefore, it appears that the women lost to follow-up did not have significantly different exposure profiles from the remaining cohort and it is very unlikely that those women lost to follow-up would differ from the remaining cohort with respect to past or recent exposure and outcome.

The only women who may have had a reasonable suspicion of their HPV status would be those women with abnormal cytology results. It is very possible that this proportion of women, with a high probability of being HPV positive, would be more likely to stay in the study so that their cervical lesions could be adequately monitored. Nonetheless, this group of women represented a small proportion of the cohort (7.4%). Furthermore, they did not differ substantially from the remaining cohort with normal cytology, with respect to exposure variables measured at baseline other than markers of sexual activity.

The use of actuarial analysis techniques to calculate the incidence or clearance rates of each group helps to ensure a more valid estimate of the risk ratio, by partitioning the total person-time experience in each risk set for the exposed and non-exposed cohorts, respectively.

9.4.3 Exposure misclassification

All members of the cohort were classified (categorized) according to specific levels or duration of exposure. Most of the exposure variables, collected from questionnaires, could vary over time from visit 1 to visit 5. Nonetheless, on the occasion that a value for a variable of interest was missing from a specific questionnaire, the response from the previous visit was carried forward. This approach could have led to misclassification of the exposure variables, since a woman may have changed her behaviour since the last visit. Re-assigning missing values with a response from a previous visit could have led to the systematic misrepresentation of the exposure distribution not only for the women who developed the outcome of interest but also for those who did not. It is very unlikely that participants who forgot to respond to a specific question were systematically more likely to also have the outcome of interest. Therefore, this misclassification is not likely to be differential. The impact of the non-differential misclassification would most likely lead to a dilution of effect (bias towards the null), but would not contribute towards the creation of spurious associations of effect.

9.4.4 Outcome misclassification

The primary outcome of interest in this study was HPV type-specific infections that had to be grouped according to their oncogenic potential for analytical purposes. While the MY09/11 PCR HPV detection system in this study is extremely sensitive and was considered state-of-the-art at the time we began the study, it has been demonstrated to be less sensitive at amplifying certain HR-HPV types in comparison to others (e.g. 35, 52, and 56). Furthermore, poor sampling of the cervix, low viral load, or presence of inhibitors in the specimen (e.g. blood) may have led to the inability to detect HPV when in fact a specimen was HPV positive. This type of misclassification could have impacted on the results in several ways. First, the occurrence of certain HPV genotypes could be underestimated in this study. Second, the overall occurrence of HR-HPV infections may be underestimated since specific HR genotypes are under-detected. Third, certain type-specific infections may have an apparently faster rate of clearance because persistent infections would be classified as cleared (under-detected). Fourth, the overall rate of HR-

HPV clearance may be underestimated if specific HR genotypes are misclassified as "cleared" because of the lower sensitivity of the PCR system to detect certain HPV types.

This potential misclassification may explain why neither HPV 35 nor HPV 52 was included in the top ten most frequent HPV genotypes detected in our study. However, there is no perfect HPV detection system and it appears that the only way to improve the sensitivity of HPV DNA detection is to utilize more than one primer system [Harnish et al., 2000] which is a very costly solution.

Despite the imperfect PCR assay, there is no reason to believe that the sample collection or type-specific sensitivity of our HPV detection system was related to the distribution of putative risk factors for HPV acquisition or clearance. Therefore, if the measures of association are biased due to the misclassification of certain outcomes, this misclassification is much more likely to be non-differential and the estimates biased towards the null.

9.4.5 Generalizability

One of the weaknesses of this cohort is that it is not a true sample of all female students from McGill or Concordia University. Attempts were made to advertise the study to students on campus, through various forms of media. Nonetheless, a systematic attempt to invite a random sample of all female students enrolled in their first or second year at university was not pursued. Instead, our cohort represents a convenient sample of students who were already attending the university health clinics, and therefore cannot necessarily be generalized to the entire female population at McGill and Concordia.

Another limitation of the study was that we were not able to obtain information on the number of women who refused to participate, or reasons why. An attempt was made to estimate what proportion of women who attended the clinic also agreed to participate in our cohort study. Unfortunately, this was a very crude estimate and we have no idea how the group of women who refused to participate differs from the cohort with respect to risk behaviour or lifestyle characteristics. Nonetheless, we were able to compare the number

of cervical abnormalities among women within our cohort with those women attending the clinic from which the cohort was drawn but who were not part of the cohort. The results of this comparison were already presented in the methods section of this thesis and showed that the proportion of cytological abnormalities was very similar in both groups.

Furthermore, this potential lack of generalizability does not affect the internal validity of our study or the estimates of risk. While the prevalence of certain risk factors may have differed between our cohort and the remaining women at McGill and Concordia, the association between the exposures of interest and the outcome would not be biased, although the precision of the point estimate will be reduced if the exposure of interest in our cohort is rarer than in the general female university population. It is noteworthy that with few exceptions, the entire body of knowledge on the epidemiology of HPV infections has been derived from studies that are not population-based, such as ours.

9.5 Future research directions

9.5.1 Genetic susceptibility for risk of persistent infections and progression to SIL

The host immune response is thought to be of critical importance in the maintenance of an HPV infection. The HLA genes, particularly the class II HLA alleles, are the primary mediators of cell-mediated immune system responses to viruses and are highly polymorphic [Maciag & Villa, 1999]. Therefore, inherited alleles may be a contributing factor in the outcome of HPV infections and cervical neoplasia. The future identification of specific high-risk HLA haplotypes for persistent HPV infections and cervical neoplasia may necessitate the re-evaluation of risk factors identified for persistent HPV infections to confirm whether or not they still explain the same level of risk according to HLA haplotype status.

9.5.2 Gene-variant interactions for risk of SIL

Viral factors may also interact with the host's genes facilitating HPV-induced cervical carcinogenesis [Ferenczy & Franco, 2002]. Results from one study observed that the risk of cervical neoplasia was substantially increased after exposure to the HPV 16 nt350G

molecular variant among women with specific P53 genetic polymorphisms [Rozendaal et al., 2000]. It would be interesting to investigate these possible interactions in more detail to help clarify if the associations that have been observed between HPV-16/18 Non-European variants and SIL would be modified according to genetic susceptibility.

9.5.3 Co-infections and risk of persistent HPV and SIL

Future studies should also evaluate the clinical significance of co-infections. The majority of women with an HPV infection in our study had a co-infection with another HPV type at one or more visits. The scope of this project prohibited extensive analysis of co-infections and associated risk of persistent HPV infections and progression to cervical neoplasia.

Young women, from one cohort study, with a co-infection with two or more HPV types at the same visit were observed to be more likely to have a same-type persistent HPV infection at the subsequent visit, and the longer the infection persisted, the harder it was to clear [Ho et al., 1998]. However, two recent studies showed that persistence of an HPV infection was independent of co-infection with other HPV types [Liaw et al., 1999; Rousseau et al., 2001]. Nonetheless, this issue needs to be further explored, in light of future HPV screening programs that are not presently designed to detect co-infections or LR-HPV types.

9.5.4 Development of a standardized approach for informing patients of an HPV infection

While randomized clinical trials are currently investigating the value of incorporating HPV testing into the triage of cervical neoplasia, some gynecologic clinics already offer HPV testing as a service. Nonetheless, it is imperative that a standardized approach for informing patients of HPV results be developed before HPV testing becomes routine. Given the high prevalence of HPV in young sexually active women, there will be an alarmingly high number of women who are informed that they have an HPV infection. From the experience in our cohort study, women were frequently upset and anxious after being informed about an HPV positive test result (at their last visit). There is still a strong

social stigma associated with an STD diagnosis and a health practitioner's attitude can play an important role in minimizing the anxiety associated with an HPV infection. Furthermore, clinicians and other health practitioners will need to understand the natural history of the infection so that they can adequately educate women about this predominantly transient infection and inform them of the co-factors that may increase the risk of HPV persistence.

9.6 Public Health Implications

This project aimed to describe the dynamics of specific HPV types, to improve the understanding of the etiology of persistent HPV infections and identify potentially important environmental determinants of HPV clearance. As a secondary objective, the goal was to investigate the risk associated with certain HPV variants and cervical lesion development.

Knowledge from this study and other similar studies of the natural history of HPV may eventually help answer questions related to viral transmissibility and the frequency of screening based on HPV testing. Furthermore, it may help provide clinicians with some guidance when counseling women with an HPV infection.

Results from this study show that HPV infections in young women attending university are very common and that there is great variability in the average duration of specific HPV type infections. Health practitioners need to be aware of the variation in duration of HPV infections and remind women that it may take as long as two-years to clear an HPV infection, but that the vast majority of infections do clear. Furthermore, while certain risk-behaviours may be very difficult to modify, others, such as using sanitary napkins instead of tampons during an HPV infection may be a more readily modifiable behaviour, with a significant impact on increasing the clearance rate of HR-HPV infections. Women who never consume dairy products may be a potentially high-risk group for HPV acquisition. However, more studies are needed to corroborate this finding. Other healthy lifestyle choices such as a diet rich in vegetables and tobacco cessation should be

emphasized. Furthermore, despite inconsistent results in the literature, condom use may still confer some protection against both HPV acquisition and persistence.

CHAPTER 10: CONCLUSIONS AND HIGHLIGHTS

1. Occurrence of HPV

Acquisition of HPV is fairly common and the two-year cumulative incidence for any HPV infection was 36% and over 50% of the HPV infections were co-infections, with two or more HPV types detected at the same visit. The average duration of infection was over one year, regardless of HPV type.

2. Determinants of incident HPV infections

Some important determinants that may facilitate the sexual transmission of an HPV infection were identified, including a recent *Chlamydia* infection and moderate alcohol consumption. Some determinants that may hinder the acquisition of an HPV infection were also identified and include washing regularly after sexual intercourse and inclusion of dairy products in the diet.

3. Predictors of clearance

Profiles for HPV clearance did not vary substantially by oncogenic HPV group, but daily consumption of vegetables and the use of sanitary napkins instead of tampons were the two strongest modifiable determinants of HR-HPV clearance.

4. HPV variants and risk of LSIL

HPV 16 or -18 variants were not associated with increased HPV persistence. However, certain Non-European variants of HPV 16 or -18 may be more strongly associated with LSIL compared to European variants.

BIBLIOGRAPHY

Abrahamowicz M, MacKenzie T, Esdaile J. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. J Am Stat Assoc 1996; 91:1432-9.

Agorastos T, Lambropoulos A, Constantinidis T, Kotsis A, Bontis J. P53 codon 72 polymorphism and risk of intra-epithelial and invasive /cervical neoplasia in Greek women. Eur J Cancer Prev 2000; 9:113-8.

Ahdieh L, Klein R, Burk R, Cu-Uvin S, Schuman P, Duerr A, Safaeian M, Astemborski J, Daniel R, Shah K. Prevalence, incidence and type-specific persistence of human papillomavirus in human immunodeficiency virus (HIV)-positive and HIV-negative women. J Infect Dis 2001; 184:682-90.

Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. Br J Cancer 1995; 72(2):511-8.

Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool . J Mol Biol 1990; 215:403-10.

Anderson-Ellstrřm A, Hagmar B, Johansson B, Kalantari M, Wärleby B, Forssman L. Human papillomavirus desoxyribonucleic acid in cervix only detected in girls after coitus. Int J STD AIDS 1996; 7:333-6.

Anttila T, Saikku P, Koskela P, Bloigu A, Dillner J, Ikaheimo I, Jellum E, Lehtinen M, Lenner P, Hakulinen T, Narvanen A, Pukkala E, Thoresen S, Youngman L, Paavonen J. Serotypes of *Chlamydia trachomatis* and risk of development of cervical squamous cell carcinoma. JAMA 2001; 285(1):47-51.

Apple R, Erlich H, Klitz W, Manos M, Becker T, Wheeler C. HLA DR-DQ associations with cervical carcinoma show papillomavirus-type specificity. Nature Gen 1994; 6:157-62.

Auborn K, Woodworth C, Dipaolo J, Bradlow H. The interaction between HPV infection and estrogen metabolism in cervical carcinogenesis. Int J Cancer 1991; 49:867-9.

Aurelian L, Altmann A, Jochmus I, Rosl F, Altmann A, Jochmus I, Rosl F. Is human papillomavirus a primary factor in the causation of cervical neoplasia? Does herpes simplex virus play a subordinate role? Con. J Gynecol Surv 1989; 5:229-34.

Banks L, Matlashewski G. Cell transformation and the HPV E5 gene. Papillomavirus Report 1993; 4:1-4.

Bauer H, Greer C, Manos M. Determination of genital human papillomavirus infection by consensus PCR amplification. In: Herrington CS MJ, editors. Diagnostic Molecular Pathology: a practical approach. Oxford: Oxford University Press, 1992.

Bauer H, Hildesheim A, Schiffman M, Glass A, Rush B, Scott D, Cadell D, Kurman R, Manos M. Determinants of genital human papillomavirus infection in low-risk women in Portland, Oregon. Sex Transm Dis 1993; 20:274-8.

Bauer H, Ting Y, Greer C, Chambers J, Tashiro C, Chimera J, Reingold A, Manos M. Genital human papillomavirus infection in female university students as determined by a PCR-based method. JAMA 1991; 265:472-7.

Beckman G, Birgander R, Sjalander A, Saha N, Holmberg P, Kivela A, Beckman L. Is p53 polymorphism maintained by natural selection? Hum Hered 1994; 44(5):266-70.

Bedell M, Jones K, Laimins L. The E6-E7 region of human papillomavirus type 18 is sufficient for transformation of NIH 3T3 and rat-I cells. J Virol 1987; 61:3635-40.

Bergman A, Nalick R. Prevalence of human papillomavirus infection in men: comparison of the partners of infected and unifected women. J Reprod Med 1992; 37:710-12.

Bertorelle R, Chieci-Bianchi L, Del Mistro A. Papillomavirus and p53 codon 72 polymorphism . Int J Cancer 1999; 82:616-17.

Berumen J, Ordonez R, Lazcano E, Salmeron J, Galvan S, Estrada R, Yunes E, Garcia-Carranca A, Gonzalez-Lira G, Madrigal de la Campa A. Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study. J Natl Cancer Inst 2001; 9(17):1325-30.

Bontkes H, van Duin M, De Gruijl T, Duggan-Keen M, Walboomers J, Stukart M, Verheijen R, Helmerhorst T, Meijer C, Scheper R, Stevens F, Dyer P, Sinnott P, Stern P. HPV 16 infection and progression of cervical intra-epithelial neoplasia: analysis of HLA polymorphism and HPV 16 E6 sequence variants. Int J Cancer 1998; 78:166-71.

Bosch F, Castellsague X, Munoz N, de Sanjose S, Ghaffari A, Gonzalez L, Gili M, Izarzugaza I, Viladiu P, Navarro C, Vergara A, Ascunce N, Guerrero E, Shah K. Male sexual behaviour and human papillomavirus DNA: key risk factors for cervical cancer in Spain. J Natl Cancer Inst 1996; 88:1060-7.

Bosch F, Lorincz A, Munoz N, Meijer C, Shah K. The causal relation betwen human papillomavirus and cervical cancer. J Clin Pathol 2002; 55:0-21.

Bosch F, Manos M, Muńoz N, Sherman M, Jansen A , Peto J, Schiffman M, Moreno V, Kurman R, Shah KatIBSoCCISG. Prevalence of human papillomavirus in cervical cancer-a worldwide perspective. J Natl Cancer Inst 1995; 87:796-802.

Bosch F, Munoz N, de Sanjose S, Guerrerro E, Chaffari A, Kaldor J, Castellsague X, Shah K. Importance of human papillomavirus endemicity in the incidence of cervical cancer - an extension of the hypothesis on sexual behavior. Cancer Epidemiol Biomarkers Prevention 1994; 3:375-9.

Bosch F, Munoz N, de Sanjose S, Izarzugaza I, Gili M, Viladiu P, Tormo MJ, Moreo P, Ascunce N, Gonzalez L, Tafur L, Kaldor J, Guerrero E, Aristizabal N, Santamaria M, Deruiz P, Shah K. Risk factors for cervical cancer in Colombia and Spain. Int J Cancer 1992; 52:750-8.

Bosch F, Rohan T, Schneider A, Frazer I, Pfister H , Castellsague X, de Sanjose S, Moreno V, Puing-Tintore L, Smith P, Munoz N, zur Hausen H. Papillomavirus research update: highlights of the Barcelona HPV 2000 international papillomavirus conference. J Clin Pathol 2001; 54:163-75.

Bosch F and de Sanjose S. Human papillomavirus and cervical cancer-burden and assessment of causality. J Natl Cancer Inst. Monographs 2003; 31:3-13.

Boyd J, Doll R. A study of the aetiology of carcinoma of the cervix uteri. Br J Cancer 1964; 18:419-34.

Brabin L. Interactions of the female hormonal environment, susceptibility to viral infections and disease progression. AIDS Patient Care 2002; 16(2):211-21.

Brinton L. Epidemiology of cervical cancer-overview. In: Munox n, Bosch FX, Shah KV, Meheus A., editors. The epidemiology of human papillomavirus and cervical cancer. Vol. 119. Lyon: IARC Scientific Publications, 1992: 3-22.

Brinton L, Hamman R, Huggins G, Lehman H, Levine R, Mallin K, Fraumeni J. Sexual and reproductive risk factors for invasive squamous cell cervical cancer. J Natl Cancer Inst 1987; 79:23-30.

Brinton L, Reeves W, Brenes M, Herrero R, Gaitan E , Tenorio F, Britton R, Garcia M, Rawls W. The male factor in the etiology of cervical cancer among sexually monogamous women. Int J Cancer 1989; 44:199-203.

Brisson J, Bairati I, Morin C, Fortier M, Bouchard C, Christen A, Bernard P, Roy M, Meisels A. Determinants of persistent detection of human papillomavirus DNA in the uterine cervix. J Infect Dis 1996; 173:794-9.

Brisson J, Morin C, Fortier M, Roy M, Bouchard C, Leclerc J, Christen A, Guimont C, Penault F, Meisels A. Risk factors for cervical intraepithelial neoplasia: differences between low and high grade lesions. Am J Epidemiol 1994; 140:700-10.

Brisson J, Roy M, Fortier M, Bouchard C, Meisels A. Condyloma and intraepithelial neoplasia of the uterine cervix: a case-control study. Am J Epidemiol 1988; 128(2):427-29.

Burk R, HO G, Beardsley L, Lempa M, Peters M, Bierman R. Sexual behavior and partner characteristics are the predominant risk factors for genital human papillomavirus infection in young women. J Infect Dis 1996; 174:679-89.

Butterworth C, Hatch K, Gore H, Mueller H, Krumdieck C. Improvement in cervical dysplasia associated with folic acid therapy in users of oral contraceptives. Am J Clin Nutr 1982; 35:72-82.

Butterworth C, Hatch K, Macaluso M, Cole P, Sauberlich H, Soong S, Borst M, Baker V. Folate deficiency and cervical dysplasia. JAMA 1992; 267:528-33.

Byers T, Marshall J, Anthony E, Fiedler R, Zielezny M. The reliability of dietary history from the distant past. Am J Epidemiol. 1987; T 125:999-1011 .

CAMH. 2000; Available at http://www.camh.net/press_releases/canadian_campus_survey_es.html.

Campo M, Moar M, Jarrett W, Laird H. A new papillomavirus associated with alimentary cancer in cattle. Nature 1980; 286:180-82.

Canadian Cancer Stats. 2000; Available at www.cancer.ca/stats

Castellsague X, Bosch FX, Munoz N. Environmental co-factors in HPV carcinogenesis. Virus Res 2002a; 89(2):191-9.

Castellsague X, Bosch F, Munoz N, Meijer C, Shah K, de Sanjose S, Eluf-Neto J, Ngelangel C, Chichareon S, Smith J, Herrero R, Moreno V, Franceschi S, International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Male circumcision, penile human papillomavirus (HPV) infection and cervical cancer. N Engl J Med 2002b; 346(15):1105-12.

Castellsague X, Ghaffari A, Daniel R, Bosch F, Munoz N, Shah K. Prevalence of penile human papillomavirus DNA in husbands of women with and without cervical neoplasia: a study in Spain and Columbia. J Infect Dis 1997; 176:353-61.

Castle P, Hillier S, Rabe L, Hildesheim R, Herrero RBM, Sherman M, Burk R, Rodriguez A , Alfaro M, Hutchinson M, Morales J, Schiffman M. An association of cervical inflammation with high-grade cervical neoplasia in women infected with oncogenic human papillomavirus (HPV). Cancer Epidemiol Biomarkers & Prev 2001; 10:1021-27.

Castle P, Schiffman M, Gravitt P, Kendall H, Fishman S, Dong H, Hildesheim A, Herrero R, Bratti M, Sherman M. Comparison of HPV DNA detection by MY09/11 PCR methods. J Med Virol 2002; 68:417-23.

Castle P and Giuliano AR. Genital tract infections, cervical inflammation and antioxidant nutrients-assessing their roles as human papillomavirus cofactors. J Natl Cancer Inst 2003; 31:29-34.

Chan P, Chang A, Cheung J, Chan D, Xu L, Tang N, Cheng A. Determinants of cervical human papillomavirus infection: Differences between high- and low-oncogenic risk types. J Infect Dis 2002; 185:28-35.

Chan S, Ho L, Ong C, Chow V, Drescher B, Duerst M, Meulen J, Villa L, Luande J, Mgaya H, Bernard H. Molecular variants of human papilllomavirus-16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. J Virol 1992; 66:2057-66.

Childers J, Chu J, Voigt L, Feigl P, Tamimi H, Franklin E, Alberts D, Meyskens F. Chemoprevention of cervical cancer with folic acid: a phase III Southwest Oncology Group intergroup study. Cancer Epidemiol Biomark Prev 1995; 4:155-59.

Chong T, Chan W, Bernard H. Transcriptional activation of human papillomavirus 16 by nuclear factor I, AP1, steroid receptors and a possibly novel transcription factor, PVF: a model for the composition of genital papillomavirus enhancers. Nucl Acids Res 1990; 18:465-70.

Clarke B, Chetty R. Postmodern cancer: the role of human immunodeficiency virus in uterine cervical cancer. J Clin Pathol 2002; 55:19-24.

Corriveau A. Cancer incidence and mortality in the NWT 1991 to 1996. Epi-North 1997; 9:5.

Coutlee F, Gravitt P, Kornegay J, Hankins C, Richardson H, Lapointe N, Voyer H, The Canadian Women's HIV study Group, Franco E. Use of PGMY primers in L1 consensus PCR improves detection of human papillomvairus DNA in genital samples. J Clin Microbiol 2002; 40:902-7.

Coutlée F, Gravitt P, Richardson H, Hankins C, Franco E, Lapointe N, Voyer H, & The Canadian women's HIV Study Group. Nonisotopic detection of typing of human papillomavirus DNA ingenital samples by the line blot assay. J Clin Microbiol 1999; 37(6):1852-7.

Cox D. Regression models and life tables (with discussion). JR Statist Soc Series B 1972; 34:187-220.

Crook T, Morgenstern J, Crawford L, Banks L. Continued expression of HPV-16 E7 protein is required for maintenance of the transformed phenotype of cells co-transformed by HPV-16 plus EJ-ras. EMBO J 1989; 8:513-19.

Cuzick J. Human papillomavirus testing for primary cervical cancer screening. JAMA 2000; 283:108-9.

Cuzick J, Terry G, Ho J, monaghan J, Lopes A, Clarkson P, Duncan I. Association between HR-HPV types, HLA DRB1* and DQB1* alleles and cervical cancer in British women. Br J Cancer 2000; 82(7):1348-52.

de Roda Husman A. Detection of human papillomavirus genotypes in normal and abnormal cervical smears Amsterdam: Vrije University of Amsterdam, 1995.

de Sanjose S, Munoz N, Bosch F, Reimann K, Pedersen N, Orfila J, Ascunce N, Gonzalez L, Tafur L, Gili M *et al.* Sexually transmitted agents and cervical neoplasia in Colombia and Spain. Int J Cancer 1994; 56:358-63.

de Villiers E. Human pathogenic papillomavirus types: an update. Curr Top Microbiol Immunol 1994; 186:1-12.

Deacon J, Evans C, Yule R, Desal M, Binns W, Taylor C, Peto J. Sexual behaviour and smoking as determinants of cervical HPV infection and of CIN3 among those infected: a case-control study nested within the Manchester cohort. Br J Cancer 2000; 88(11):1565-72.

Dyson N, Howley P, Munger K, Harlow E. The human papillomavirus 16 E7 oncoprotein is able to bindto the retinoblastoma gene product. Science 1989; 243:934-37.

Elfgren K, Kalantari M, Moberger B, Hagmar B, Dillner J. A population-based five-year follow-up study of cervical human papillomavirus infection. Am J Obstet Gynecol 2000; 183(3):561-7.

Ellis J, Keating P, Baird J, Hounsell E, Renouf D, Rowe M, Hopkins D, Duggan-Keen M, Bartholomew J, Young L *et al.* The association of an HPV 16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. Nat Med 1995; 5:464-70.

Eto I, Krumdieck C. Role of vitamin B12 and folate deficiencies in carcinogenesis. Adv Exp Med Biol 1986; 206:313-30.

Evander M, Edlund K, Gustafsson A, Jonsson M, Karlsson R, Rylander E, Wadell G. Human papillomavirus infection is transient in young women: A population based cohort study. J Infect Dis 1995; 171:1026-30.

Fairley C, Chen S, Tabrizi S, Leeton K, Quinn M, Garland S. The absence of genital human papillomavirus DNA in virginal women. Int J Sex Transm Dis AIDS 1992; 3:414-7.

Ferenczy A, Bergeron C, Richart R. Human papillomavirus DNA in fomites on objects used for the management of patients with genital human papillomavirus infections. Obstet Gynecol 1989; 74:950-4.

Ferenczy A, Bergeron C, Richart R. Human papillomavirus infection in CO2 laser-generated plume of smoke and its consequences to the surgeon. Obstet Gynecol 1990; 75:114-8.

Ferenczy A, Franco E. Persistent human papillomavirus infection and cervical neoplasia. Lancet 2002; 3:11-6.

Ferlay J, Parkin D, Pisani P. GLOBOCAN 1: cancer incidence and mortality worldwide. IARC CancerBase no 3. Lyon: IARCPress, 1998.

Franco E. The sexually transmitted disease model for cervical cancer: incoherent epidemiologic findings and the role of misclassification of human papillomavirus infection. Epidemiology 1991a; 2:98-106.

Franco E. Viral etiology of cervical cancer: a critique of the evidence. Rev Infect Dis 1991b; 13:1195-206.

Franco E. Measurement errors in epidemiological studies of human papillomavirus and cervical cancer. In: Munoz N, Bosch FX, Shah KV, Meheus A., (eds.). The epidemiology of human papillomavirus and cervical cancer. Lyon: WHO-IARC, 1992: 181-97.

Franco E. Human papillomavirus and the natural history of cervical cancer. Infect Med 1993; 10:57-64.

Franco E, Duarte-Franco E, Ferenczy A. Cervical cancer: epidemiology, prevention, and the role of human papillomavirus infection. CMAJ 2001; 164(7):1017-25.

Franco E, Villa L, Rahal P, Ruiz A. Molecular variant analysis as an epidemiological tool to study persistence of cervical human papillomavirus infection. J Natl Cancer Inst 1994; 86:1557-8.

Franco E, Villa L, Ruiz A, Costa M. Transmission of cervical human papillomavirus infection by sexual activity: differences between low and high oncogenic risk types. J Infect Dis 1995; 172:756-63.

Franco E, Villa L, Sobrinho J, Prado J, Rousseau M , Desy M, Rohan T. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. J Infect Dis 1999; 180(5):1415-23.

Gagnon F. Contribution to the study of the etiology and prevention of cancer of the cervix of the uterus. Am J Obstet Gynecol 1950; 60:516-22.

Gallego M, Zimonjic D, Popescu N, Dipaolo J, Lazo P. Integration site of human papillomavirus type-18 DNA in chromosome band 8q22.1 of C4-1 cervical carcinoma: DNase I hypersensitivity and methylation of cellular flanking sequences. Genes Chromosomes Cancer 1994; 9:28-32.

Gaudette L, Altmayer C, Wysocki M, Gao R. Cancer incidence and mortality across Canada. Health Rep 1998; 10:51-66.

Giannoudis A, Herrington C. Human papillomavirus variants and squamous neoplasia of the cervix. J Pathol 2001; 193:295-302.

Giuliano A, Harris R, Sedjo R, Baldwin S, Roe D, Papenfuss M, Abrahamsen M, Inserra P, Olvera S, Hatch K. Incidence, prevalence, and clearance of type specific human papillomavirus infections: The young women's health study. J Infect Dis 2002a; 186:462-9.

Giuliano A, Papenfuss M, Abrahamsen M, Inserra P. Differences in factors associated with oncogenic and cononcogenic human papillomavirus infection at the United States-Mexico border. Cancer Epidemiol Biomarkers & Prev 2002b; 11:930-4.

Giuliano A, Papenfuss M, Nour M, Canfield L, Schneider A, Hatch K. Antioxidant nutrients: associations with persistent human papillomavirus infection. Cancer Epidemiol Biomarkers & Prev 1997; 6:917-23.

Giuliano A, Papenfuss M, Schneider A, Nour M, Hatch K. Risk factors for high-risk type human papillomavirus infection among Mexican-American women. Cancer Epidemiol, Biomarkers Prev 1999; 8:615-20.

Giuliano A, Sedjo R, Roe D, Harri R, Baldwin S, Papenfuss M, Abrahamsen M, Inserra P. Clearance of oncogenic human papillomavirus (HPV) infection: effect of smoking (United States). Cancer Causes Control 2002; 13(9):839-46.

Goldberg G, Vermund S, Schiffman M, Ritter D, Spitzer C, Burk R. Comparison of cytobrush and cervicovaginal lavage sampling methods for the detection of genital human papillomavirus. Am J Obstet Gynecol 1989; 161 :1669-72.

Gravitt P, Peyton C, Alessi T, Wheeler C, Coutlée F, Hildesheim A, Schiffman M, Scott D, Apple R. Improved amplification of genital human papillomaviruses. J Clin Microbiol 2000; 38(1):357-61.

Gravitt P, Peyton C, Apple R, Wheeler C. Genotyping of 27 human papillomavirus types by using L1 consensus PCR products by single-hybridization, reverse line blot detection method. J Clin Microbiol 1998; 36:3020-7.

Greer C, Resnick R, Peterson S *et al.* PCR-based methods for the analysis of HPV in paraffin-embedded tissues and the effect of fixation methods and age. Papillomavirus Workshop. 1990: 26.

Gustafsson L, Ponten J, Zack M, Adami H. International incidence rates of invasive cervical cancer after introduction of cytological screening. Cancer Causes Control 1997; 8:755-63.

Hamel P, Gallie B, Phillips R. The retinoblastoma protein and cell cycle regulation. Trends Genet 1992; 8:180-5.

Hankinson S, Willett W, Manson J, Hunter D, Coldizt G, Stampfer M, Longcope C, Speizer F. Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal women. J Natl Cancer Inst 1995; 87:1297-302.

Harnish D, Belland L, Scheid E, Rohan T. Evaluation of human papillomavirus consensus primers for HPV detection by the polymerase chain reaction. Molecular & Cellular Probes 2000; 13:9-21.

Harro C, Yuk-Ying S, Roden R, Hildesheim A, Wang Z, Reynolds M, Mast T, Robinson R, Murphy B, Karron R, Dillner J, Schiller J, Lowy D. Safety and immunogenicity trial in adut volunteers of a human papillomavirus 16 L1 virus-like particle vaccine. J Natl Cancer Inst 2001; 93(4):284-92.

Healey S, Aronson K, Mao Y, Schlecht N, Mery L, Ferenczy A, Franco E. Oncogenic human papillomavirus infection and cervical lesions in aboriginal women of Nunavut, Canada. Sex Transm Dis 2001; 28(12):694-700.

Herrero R, Hildesheim A, Bratti C, Sherman M, Hutchinson M, Morales J, Balmaceda I, Greenberg M, Alfaro M, Burk R, Wacholder S , Plummer M, Schiffman M. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. J Natl Cancer Inst 2000; 92(6):464-74.

Hildesheim A, Gravitt P, Schiffman M, Kurman R, Barnes W, Jones S, Tchabo J, Brinton L, Copeland C, Epp J, Manos M. Determinants of genital human papillomavirus infection in low-income women in Washington, D.C. Sex Transm Dis 1993; 20:279-85.

Hildesheim A, Herrero R, Castle P, Wacholder S, Bratti M, Sherman M, Lorincz A, Burk R, Morales J, Rodriguez A, Helgesen K, Alfaro M, Hutchinson M, Balmaceda I, Greenberg M, Schiffman M. HPV co-factors related to the development of cervical cancer: results from a population-based study in Costa Rica. Br J Cancer 2001a; 84(9):1219-26.

Hildesheim A, Schiffman M, Bromley C, Wacholder S, Herrero R, Rodriguez C, Bratti M, Sherman M , Scarpidis U, Lin Q, Terai M, Bromley R, Buetow K , Apple R, Burk R. Human papillomavirus type 16 variants and risk of cervical cancer. J Natl Cancer Inst 2001b; 93(4):315-8.

Hildesheim A, Schiffman M, Gravitt P, Glass A, Greer C, Zhang T, Scott D, Rush B, Lawler P, Sherman M, Kurman R, Manes M. Persistence of type-specific human papillomavirus infection among cytologically normal women. J Infect Dis 1994; 169:235-40.

Hildesheim A, Schiffman M, Scott D, Marti D, Kissner T, Sherman M, Glass A, Manos M, Lorincz A, Kurman R, Buckland J, Rush B, Carrington M. Human leukocyte antigen class I/II alleles and development of human papillomavirus-related cervical neoplasia: Results from a case-control study conducted in the United States. Cancer Epidemiol Biomarkers & Prev 1998; 7:1035-41.

Ho G, Bierman R, Beardsley L, Chang C, Burk R. Natural history of cervicovaginal papillomavirus infection in young women. N Engl J Med 1998; 338(7):423-8.

Ho G, Burk R, Fleming I, Klein R. Risk for human papillomavirus infection in women with human immunodeficiency virus-induced immunosuppression. Int J Cancer 1994; 56:788-92.

Ho G, Burk R, Klein S, Kadish A, Chang C, Palan P, Basu J, Tachezy R, Lewis R, Romney S. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. J Natl Cancer Inst 1995; 87(18):1365-71.

Ho G, Kadish A, Burk R, Basu J, Pala P, Mikhail M, Romney S. HPV 16 and cigarette smoking as risk factors for high-grade cervical intra-epithelial neoplasia. Int J Cancer 1998; 78:281-85.

Ho L, Chan S, Burk R, Das B, Fujinaga K, Icenogle J, Kahn T, Kiviat N, Lancaster W, Mavromara-Nazos P *et al.* The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. J Virol 1993; 67:6413-23.

Ho LCSCVCTTSVLBH. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. J Clin Microbiol 1991; 29:1765-72.

Holford T. Multivariate methods in epidemiology. (Monographs in epidemiology and biostatistics; 32). New York: Oxford University Press, Inc., 2002.

Holowaty P, Miller A, Rohan T, To T. Natural history of dysplasia of the uterine cervix. J Natl Cancer Inst 1999; 91(3):252-58.

IARC. IARC Monographs on the evaluation of carcinogenic risks to humans. Volume 64. Human Papillomaviruses. Lyon, France: IARC Scientific Publications , 1995.

Jenison S, Yu X, Valentine J, Koutsky L, Christiansen A, Beckmann A, Galloway D. Evidence of prevalent genital-type human papillomavirus infections in adults and in children. J Infect Dis 1990; 162:60-9.

Johnson K, Canadian Task Force on the Periodic Health Examination. Periodic Health Examination, 1995 update: 1. Screening for human papillomavirus infection in asymptomatic women. CMAJ 1995; 152(4):483-93.

Jones E, MacDonald I, Breslow L. A study of epidemiologic factors in carcinoma of the uterine cervix. Am J Obstet Gynecol 1958; 76:1-10.

Joseffson A, Magnusson P, Ylitalo N, Quarforth-Tubbin P, Ponten J, Adami H, Gyllensten U. p53 polymorphism and risk of cervical cancer. Nature (Lond.) 1998; 396:531.

Juarez-Figueroa L, Wheeler C, Uribe-Salas F, Conde-Glez C, Zampilpa-Mejia L, Garcia-Cisneros S , Hernandez-Avila M. Human papillomavirus: a highly prevalent sexually transmitted disease agent among female sex workers from Mexico City. Sex Transm Dis 2001; 28:125-30.

Kalbfleisch J, Prentice RL. The statistical analysis of failure time data. New York: Wiley, 1980.

Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Amer Statist Assoc 1958; 53:457-81.

Kaye J, Starkey W, Kell B, Biswas C, Raju K, Best J, Cason J. Human papillomavirus type 16 in infants: use of DNA sequence analyses to determine the source of infection. J Gen Virol 1996; 77:1139-43.

Kjaer S. Risk factors for cervical neoplasia in Denmark. APMIS 1998; 80:1-41.

Kjaer S, De Villiers E, Caglayan H, Svare E, Haugaard B, Engholm G, Christensen R, Moller K, Poll P, Jensen H, Vestergaard B, Lynge E, Jensen O. Human papillomavirus, Herpes-simplex virus and other potential risk factors for cervical cancer in a high-risk area (Greenland) and a low-risk area (Denmark): A second look. Br J Cancer 1993; 67:830-7.

Kjaer S, de Villiers E, Dahl C, Engholm G, Bock J, Vestergaard B, Lynge E, Jensen O. Case-control study of risk factors for cervical neoplasia in Denmark. I: Role of the "male factor" in women with one lifetime sexual partner. Int J Cancer 1991; 48:39-44.

Kjaer S, Engholm G, Teisen C, Haugaard B, Lynge E , Christensen R, Moller K, Jensen H, Poll P, Vestergaard B, DeVilliers E, Jensen O. Risk factors for cervical human papillomavirus and herpes simplex virus infections in Greenland and Denmark: a population-based study. Am J Epidemiol 1990; 131:669-82.

Kjaer S, van den Brule A, Boch J, Poll P, Engholm G, Sherman M, Walboomers J, Meijer C. Human papillomavirus-the most significant risk determinant of cervical intraepithelial neoplasia. Int J Cancer 1996; 65:601-6.

Kjaer S, van den Brule A, Bock J, Poll P, Engholm G, Sherman M, Walboomers J, Meijer C. Determinants for genital human papillomavirus infectionn in 1000 randomly chosen young Danish women with normal Pap smear. Different risk profiles for oncogenic and non-oncogenic HPV types? Cancer Epidemiol Biomarkers & Prevention 1997; 6:799-805.

Kjellberg L, Hallmans G, Ahren A, Johansson R, Bergman F, Wadell G, Angstrom T, Dillner J. Smoking, diet, pregnancy and oral contraceptive use as risk factors for cervical intra-epithelial neoplasia in relation to human papillomavirus infection. Br J Cancer 2000; 82(7):1332-38.

Klaes R, Ridder R, Schaeffer U, Benner A, von Knebel Doeberitz M. No evidence of p53 allele-specific predisposition in human papillomavirus-associated cervical cancer. J Mol Med 1999; 77:299-302.

Kleinbaum D. Survival Analysis. New-York: Springer-Verlag, 1996.

Koss L. The Papanicolaou test for cervical cancer detection. A triumph and a tragedy. JAMA 1989; 261:737-43.

Kotloff K, Wasserman S, Russ K, Shapiro S, Daniel R , Brown W, Frost A, Tabara S, Shah K. Detection of genital human papillomavirus and associated cytological abnormalities among college women. Sex Transm Dis. 1998; 25:243-50.

Koutsky L, Ault K, Wheeler C, Brown D, Barr E, Alvarez F, Chiacchierini L, Jansen K, For The Proof Of Principle Study Investigators. A controlled trial of a human papillomavirus type 16 vaccine. New Engl J Med 2002; 347(21):1645-51.

Koutsky L, Galloway D, Holmes K. Epidemiology of genital human papillomavirus infection. Epidemiol Rev 1988; 10:122-63.

Koutsky L, Holmes K, Critchlow C, Stevens C, Paavonen J, Beckmann A, DeRouen T, Galloway D, Vernon D, Kiviat N. A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. 1992; 327:1272-8.

Kruger-Kjaer S, van den Brule A, Svare E, Engholm G, Sherman M, Poll P, Walboomers J, Bock J, Meijer C. Different risk factor patterens for high-grade and low-grade intraepithelial lesions on the cervix among HPV positive and HPV negative young women. Int J Cancer 1998; 76:613-19.

Kruger-Kjaer S, Chackerian B, van den Brule A, Svare E, Paull G, Walboomers J, Schiller J, Bock J, Sherman M, Lowry D, Meijer C. High-risk human papillomavirus is sexually transmitted: evidence from a follow-up study of virgins starting sexual activity (intercourse) . Cancer Epidemiol, Biomarkers Prev 2001; 10:101-6.

La Vecchia C, Franceschi S, Decarli A, Fasoli M, Gentile A, Parazzini F, Regallo M. Sexual factors, venereal diseases and the risk of intraepithelial and invasive cervical neoplasia. Cancer 1986; 58:935-44.

Lacey J, Brinton L, Abbas F, Barnes W, Gravitt P , Greenberg M, Greene S,

Hadjimichael O, McGowan L, Mortel R, Schwartz P, Silverberg S, Hildesheim A. Oral contraceptives as risk factors for cervical adenocarcinomas and squamous cell carcinomas. Cancer Epidemiol, Biomarkers & Prev 1999; 8:1079-85.

Lane D. p53, guardian of the genome. Nature 1992; 358:15-6.

LaVecchia C, Tavani A, Franceschi S, Parazzini F. Oral contraceptives and cancer. A review of the evidence. Drug Saf. 1996; 14:260-72.

Lazcano-Ponce E, Herrero R, Muńoz N, Cruz A, Shah K, Alonso P, Hernández P, Salmerón J , Hernández M. Epidemiology of HPV infection among Mexican women with normal cervical cytology. Int J Cancer 2001; 91:412-20.

Ley C, Bauer H, Reingold A, Schiffman M, Chambers J, Tashiro C, Manos M. Determinants of genital human papillomavirus infection in young women. J Natl Cancer Inst 1991; 83:997-1003.

Liaw K, Glass A, Manos M, Greer C, Scott D, Sherman M, Burk R, Kurman R, Wacholder S, Rush B , Cadell D, Lawler P, Tabor D, Schiffman M. Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions. J Natl Cancer Inst 1999; 91(11):954-60.

Liaw K, Hildesheim A, Burk R, Gravitt P, Wacholder S, Manos M, Scott D, Sherman M, Kurman R, Glass A, Anderson S, Schiffman M. A prospective study of human papillomavirus type 16 DNA detection by polymerase chain reaction and its association with acquisition and persisitence of other HPV types. J Infect Dis 2001; 183:8-15.

Londesborough P, Ho L, Terry G, Cuzick J, Wheeler C , Singer A. Human papillomavirus genotype as a predictor of persistence and development of high-grade lesions in women with minor cervical abnormalities. Int J Cancer 1996; 69:364-8.

Lorincz A, Castle P, Sherman M, Scott D, Glass A, Wacholder S, Rush B, Gravitt P, Schussler J, Schiffman M. Viral load of human papillomavirus and risk of CIN3 or cervical cancer. Lancet 2002; 360:228-9.

Ludicke F, Stalberg A, Vassilakos P, Major A, Campana A. High- and intermediate-risk human papillomavirus infection in sexually active adolescent females. J Pediatr Adolesc Gynecol 2001; 14:171-74.

Maciag P, Schlecht N, Souza P, Franco E, Villa L, Petzl-Erler M. Major histocompatibility complex class II polymorphisms and risk of cervical cancer and human papillomavirus infection in Brazilian women. Cancer Epidemiol, Biomarkers Prev 2000; 9:1183-91.

Maciag P, Villa L. Genetic susceptibility to HPV infection and cervical cancer. Braz J Med Biol Res 1999; 32(7):915-22.

Makni H, Franco E, Kaiano J, Villa L, Labrecque S, Dudley R, Storey A, Matlashewski G. p53 polymorphism in codon 72 and risk of human papillomavirus-induced cervical cancer: effect of inter-laboratory variation. Int J Cancer 2000; 87:528-33.

Manos M, Kinney W, Hurley L, Sherman M, Shieh-Ngai J, Kurman R, Ransley J, Fetterman B, Hartinger J, McIntosh K, Pawlick G, Hiatt R. Identifying women with cervical neoplasia: Using human papillomavirus DNA testing for equivocal papanicolaou results. JAMA 1999; 281(17):1605-10.

Manos M, Ting Y, Wright D, Lewis A, Broker T, Wolinski S. Use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. Cancer Cells 1989; 7:209-14.

Martin C. Marital and copital factors in cervical cancer. Am J Public Health 1967; 57:803-14.

Mason J, Levesque T. Folate: effects on carcinogenesis and the potential for cancer chemoprevention. Oncology 1996; 10:1727-44.

Matsukura T, Sugase M. Relationships between 80 human papillomavirus genotypes and different grades of cervical intraepithelial neoplasia: association and causality. Virology 2001; 283:139-47.

Mayrand M, Coutlee F, Hankins C, Lapointe N, Forest P, de Ladurantaye M, The Candian Women's HIV Study Group, Roger M. Detection of human papillomavirus type 16 DNA in consecutive genital samples does not always represent persistent infection as determined by molecular variant analysis. J Clin Microbiol 2000; 38(9):3388-93.

McCrory D, Matchar D, Bastian L *et al.* Evaluation of cervical cytology: Evidence Report/Technology Assessment No. 5. AHCPR Publication No. 99-E010. Rockville, MD.: US Department of Health and Human Services, 1999.

McNicol P, Young T. Prevalence and correlates of human papillomavirus infection in a Canadian inner city population determined by PCR. 14th International Papillomavirus Conference. Quebec City: Faculty of Medicine, Laval University.

Meijer C, Snijders P, van den Brule A. Screening for cervical cancer: should we test for infection with high-risk HPV? CMAJ 2000; 163(5):535-8.

Melikian A, Wang X, Waggoner S, Hoffman D, El-Bayoumy K. Comparative response of normal and of human papillomavirus-16 immortalized human epithelial cervical cells to benzo[a]pyrene. Oncol Rep 1999; 6(6):1371-6.

Melkert P, Hopman E, van den Brule A, Risse E, Van Diest P, Bleker O, Helmerhorst T, Schipper M, Meijer C, Walboomers J. Prevalence of HPV in cytomorphologically normal cervical smears as determined by the polymerase chain reaction is age-dependent. Int J Cancer 1993; 53:919-23.

Miller A, Anderson G, Brisson J, Laidlaw J, Le Pitre N, Malcolmson P, Mirwaldt P, Stuart G, Sullivan W. Report of a National Workshop on Screening for Cancer of the Cervix. CMAJ 1991; 145:1301-25.

Miller A, Nazeer S, Fonn S, Brandup-Lukanow A, Rehman R, Cronje H, Sankaranarayanan R, Koroltchouk V, Syrjanen K, Singer A, Onsrud M. Report on consensus conference on cervical cancer screening and management. Int J Cancer 2000; 86:440-7.

Mittal R, Tsutsumi K, Pater A, Pater M. Human papillomavirus type 16 expression in cervical keratinocytes: role of progesterone and glucocorticoid hormones. Obstet Gynecol 1993; 81(1):5-12.

Moreno V, Bosch X, Munoz N, Meijer C, Shah K, Walboomers J, Herrero R, Franceschi S, for the IARC Multicentric Cervical Cancer Study Group. Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case-control study. Lancet 2002; 359:1085-92.

Morrison E, Goldberg G, Hagan R, Kadish A, Burk R. Self-administered home cervicovaginal lavage: a novel tool for the clinical-epidemiologic investigation of genital human papillomavirus infections. Am J Obstet Gynecol 1992; 167:104-7.

Moscicki A, Hills N, Shiboski S, Powell K, Jay N, Hanson E, Miller S, Clayton L, Farhat S, Broering J, Darragh T, Palefsky J. Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. JAMA 2001; 285(23):2995-3002.

Moscicki A, Palefsky J, Smith G, Siboshski S, Schoolnik G. Variability of human papillomavirus DNA testing in a longitudinal cohort of young women. Obstet Gynecol 1993; 82:578-85.

Moscicki A, Shiboski S, Broering J, Powell K, Clayton L, Jay N, Darragh T, Brescia R, Kanowitz S, Miller S, Stone J, Hanson E, Palefsky J. The natural history of human papillomavirus infection as measured by repeated DNA testing in adolescent and young women . Journal of Pediatrics 1998; 132(2):277-84.

Munger K, Werness B, Dyson N, Phelps W, Howley P. Complex formationof human papillomavirus E7 proteins with the retinoblastoma tumour suppressor gene product. EMBO J 1989; 8:4099-105.

Munoz N, Bosch F, Chichareon S *et al.* A multinational case-control study on the risk of cervical cancer linked to 25 HPV types:which are the high-risk types? 18th International Papillomavirus Conference. Barcelona, Spain: 2000.

Munoz N, Bosch F, de Sanjose S, Herrero R, Castellsague X, Shah K, Snijders P, Meijer C, International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 2003; 348(6):518-27.

Munoz N, Bosch F, de Sanjose S, Tafur L, Izarzugaza I, Gili M, Viladiu P, Navarro C, Martos C, Ascunce N, Gonzalez L, Kaldor J, Guerrero E, Lorincz A, Santamaria M, Deruiz P, Aristizabal N, Shah K. The causal link between human papillomavirus and invasive cervical cancer: A population-based case-control study in Colombia and Spain. Int J Cancer 1992; 52:743-9.

Munoz N, Bosch F, Desanjose S, Vergara A, Delmoral A, Munoz M, Tafur L, Gili M, Izarzugaza I, Viladiu P, Navarro C, Deruiz P, Aristizabal N, Santamaria M, Orfila J, Daniel R, Guerrero E, Shah K. Risk factors for cervical intraepithelial neoplasia grade-3 carcinoma insitu in Spain and Colombia. Cancer Epidemiol Biomarkers & Prevention 1993; 2:423-31.

Munoz N, Franceschi S, Bosetti C, Morreno V, Herrero R, Smith J *et al.* Role of parity and human papillomavirus in cervical cancer: the IARC multicentric case-control study. Lancet 2002; 359:1093-101.

Myers G, Bernard H, Delius H, Baker C, Icenogle J , Halpern A, Wheeler C. Human papillomavirus. A compilation of analysis of nucleic acid and amino acid sequences. Los Alamos, N.M 1995.

Nobbenhuis M, Helmerhorst T, van den Brule A, Rozendaal L, Voorhorst F, Bezemer P, Verheijen R, Meijer C. Cytological regression and clearance of high-risk human papillomavirus in women with an abnormal cervical smear. Lancet 2001; 358:1782-3.

Nobbenhuis M, Walboomers J, Helmerhorst T, Rozendaal L, Remmink A, Risse E, van der Linden H, Voorhorst F, Kenemans P, Meijer C. Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screeening: a prospective study. Lancet 1999; 354:20-5.

Odunsi K, Terry G, Ho L, Bell J, Cuzick J, Ganesan T. Susceptibility to human papillomavirus-associated cervical intra-epithelial neoplasia is determined by specific HLA DR-DQ alleles. Int J Cancer 1996; 67:595-602.

Odunsi K, Ganesan T. The roles of the human major histocompatibility complex and human papillomavirus infection in cervical intraepithelial neoplasia and cervical cancer. Clin Oncol 1997; 9:4-13.

Ong C, Chan S, Campo M, Fujinaga K, Mavromara-Nazos P, Labropoulou V, Pfister H, Tay S, ter Meulen J, Villa L, Bernard H. Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratypic diversity reflect co-evolution with human ethnic groups. J Virol 1993; 67:6424-31.

Pakarian F, Kaye J, Cason J, Kell B, Jewers R, Derias N, Raju K, Best J. Cancer associated human papillomavirus: perinatal transmission and persistence. Br J Obstet Gynecol 1994; 101:514-17.

Palan P, Chang C, Mikhail M, Ho G, Basu J, Romney S. Plasma concentrations of micronutrients during a nine-month clinical trial of beta-carotene in women with precursor cervical cancer lesions. Nutr Cancer 1998; 30:46-52.

Palefsky J, Holly E. Molecular virology and epidemiology of human papillomavirus and cervical cancer. Cancer Epidemiol, Biomarkers & Prev 1995; 4:415-28.

Papanicolaou G. Race Betterment Foundation New Cancer Diagnosis . . Bat-Creek, Mich: 1954.

Parkin D, Pisani P, Ferlay J. Estimates of the worldwide incidence of 25 major cancers in 1990. Int J Cancer 1999; 80:827-41.

Pater A, Bayatpour M, Pater M. Oncogenic transformation by human papillomavirus type 16 deoxynucleic acid in the presence of progesterone or progestins from oral contraceptives. Am J Obstet Gynecol 1990; 162:1099-103.

Pater M, Mittal R, Pater A. Role of steroid hormones in potentiating transformation of cervical cells by human papillomaviruses. Trends Microbiol 1994; 2(7):229-34.

Peto R, Pike M, Armitage P, Breslow N, Cox D, Howard S, Mantel N, McPherson K, Peto J, Smith P. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer 1976; 34:585-612.

Popescu N, Amsbaugh S, DiPaolo J. Integration sites of human papillomavirus 18 DNA sequences on HeLa cell chromosomes. Cytogenet Cell Genet 1987; 44:58-62.

Poppe W, Ide P, Drijkoningen M, Lauweryns J, Van Assche F. Tobacco smoking impairs the local immunosurveillance in the uterine cervix. An immunohistochemical study. Gynecol Obstet Invest 1995; 39(1):34-8.

Pridan H, Lilienfeld A. Carcinoma of the cervix in Jewish women in Israel, 1960-67: An epidemiological study. Isr J Med Sci 1971; 7:1465-70.

Quantin C, Abrahamowicz M, Moreau T, Bartlett G, MacKenzie T, Tazi M, Lalonde L, Faivre J. Variation over time ofv the effects of prognostic factors in a population-based study of colon cancer: comparison of statistical models. Am J Epidemiol 1999; 150(11):1188-200.

Ramsay J. Monotone regression splines in action, (with discussion). Stat Sci 1988; 3:425-61.

Rando R, Lindheim S, Hasty L, Sedlacek T, Woodland M, Eder C. Increased frequency of detection of human papillomavirus deoxyribonucleic acid in exfoliated cervical cells during pregnancy. Am J Obstet Gynecol 1989; 161:50-5.

Ratnam S, Franco E, Ferenczy A. Human papillomavirus testing for primary screening of cervical cncer precursors. Cancer Epidemiol, Biomarkers Prev 2000; 9:945-51.

Reeves W, Rawls W, Brinton L. Epidemiology of genital papillomaviruses and cervical cancer. Rev Infect Dis 1989; 11:426-39.

Reichman M, Judd J, Longcope C, Schatzkin A, Clevidence B, Nair P, Campbell W, Taylor P. Effects of alcohol consumption on plasma and urinary hormone concentration in premenopausal women. J Natl Cancer Inst 1993; 85:722-27.

Richardson H, Franco E, Pintos J, Bergeron J, Arella M, Tellier P. Determinants of Low- and High-Risk Cervical HPV Infections in Montreal University Students . Sex Transm Dis 2000; 27:79-86.

Richardson H, Kelsall G, Tellier P, Voyer H, Abrahamowicz M, Ferenczy A, Coutlée F, Franco E. The natural history of type-specific human papillomavirus infections in female university students. Cancer Epidemiol Biomarkers & Prev 2003;11:485-90.

Richart R. Natural history of cervical intraepithelial neoplsia. Clin Obstet Gynecol 1968; 10:748-84.

Roden R, Hubbert N, Kirnbauer R, Christensen N, Lowy D, Schiller J. Assessment of the serological relatedness of genital human papillomaviruses by hemagglutination inhibition. J Virol 1996; 70:3298-301.

Roden R, Yutzy W, Fallon R, Inglis S, Lowy D, Schiller J. Minor capsid protein of human genital papillomaviruses contains subdominant, cross-neutralizing epitopes. Virology 2000; 270:254-7.

Roman A, Fife K. Human papillomavirus DNA associated with foreskins of normal newborns. J Infect Dis 1986; 153:855-61.

Rosenthal A, Ryan A, Al-Jehani R, Storey A, Harwood C, Jacobs I. p53 codon 72 polymorphism and risk of cervical cancer in the UK. Lancet 1998; 352:871-72.

Rothman KJ, Greenland S. Modern epidemiology. Second Edition edition. Philadelphia, PA, USA: Lippincott-Raven, 1998.

Rotkin I. Adolescent coitus and cervical cancer. Cancer Res. 1967; 27:603-17.

Rousseau MC. The epidemiology of cervical coinfection with multiple human papillomavirus types in a cohort of Brazilian women [Ph.D. thesis]. Montréal, Qc, Canada: McGill University, 2003.

Rousseau M, Franco E, Villa L, Sobrinho J, Termini L, Prado J, Rohan T. A cumulative case-control study of risk factor profiles for oncogenic and nononcogenic cervical human papillomavirus infections. Cancer Epidemiol, Biomarkers Prev 2000; 9:469-76.

Rousseau M, Pereira J, Prado J, Villa L, Rohan T, Franco E. Cervical coinfection with human papillomavirus (HPV) types as a predictor of acquisition and persistence of HPV infection. J Infect Dis 2001; 184(12):1508-17.

Rozendaal L, Walboomers J, Van der Linden J, Voorhorst F, Kenemans P, Helmerhorst T, van Ballegooijen M, Meijer C. PCR-based high-risk HPV test in cervical cancer screening gives objective risk assessment of women with cytomorphological normal cevical smears. Int J Cancer 1996; 68:766-9.

Rozendaal L, Westerga J, van der Linden J, Walboomers J, Voorhorst F, Risse E, Boon M, Meijer C. PCR based high-risk HPV testing is superior to neural network based screening for predicting incident CINIII in women with normal cytology and borderline changes. J Clin Pathol 2000; 53:606-11.

Rylander E, Ruusuvaara L, Wiksten M, Evander M, Wadell G. The absence of vaginal human papillomavirus 16 DNA in women who have not experienced sexual intercourse. Obstet Gynecol 1994; 83:735-7.

Saiki R, Gelfand D, Stoffel S. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 1988; 239:487-91.

Sanjeevi C, Hjelmstrom P, Hallmans G, Wiklund F, Lenner P, Angstrom T, Dillner J, Lernmark A. Different HLA-DR-DQ haplotypes are associated with cervical intraepithelial neoplasia among human papillomavirus type-16 seropositive and seronegative Swedish women. Int J Cancer 1996; 68:409-14.

Sasagawa T, Basha W, Yamazaki H, Inoue M. High-risk and multiple human papillomavirus infections associated with cervical abnormalities in Japanese women. Cancer Epidemiol, Biomarkers Prev 2001; 10:45-52.

Schiff M, Becker T, Masuk M, van Asselt-King L, Wheeler C, Altobelli K, North C, Nahmias A. Risk factors for cervical intraepithelial neoplasia in Southwestern American Indian women. Am J Epidemiol 2000; 152(8):716-26.

Schiffman M. Recent progress in defining the epidemiology of human Papillomavirus infection and cervical neoplasia. J Natl Cancer Inst 1992; 84:394-8.

Schiffman M, Bauer H, Hoover R, Glass A, Cadell D, Rush B, Scott D, Sherman M, Kurman R, Wacholder S, Stanton C, Manos M. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J Natl Cancer Inst 1993; 85:958-64.

Schiffman M, Brinton L. The epidemiology of cervical carcinogenesis. Cancer 1995; 76:1888-901.

Schiffman M, Herrero R, Hildesheim A, Sherman M, Bratti M, Wacholder S, Alfaro M, Hutchinson M, Morales J, Greenberg M, Lorincz A. HPV DNA testing in cervical cancer screening: results from women in a high-risk province of Costa Rica. JAMA 2000; 283:87-93.

Schiffman M, Schatzkin A. Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. Cancer Res 1994; 54:S1944-S1947.

Schiffman M. and Kruger-Kjaer S. Natural history of anogenital human papillomavirus infection and neoplasia. J Natl Cancer Inst Monographs 2003; 31:14-19.

Schlecht N, Kulaga S, Robitaille J, Ferreira S, Santos M, Miyamura R, Duarte-Franco E, Rohan T, Ferenczy A, Villa L, Franco E. Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia. JAMA 2001; 286(24):3106-14.

Schlect N, Franco EL, Rohan TE, Kjaer SK, Schiffman MH, Moscicki A-B and Duffy SW, For the MARSH International Collaborative Group. Repeatability of sexual history in longitudinal studies on HPV infection and cervical neoplasia: determinants of reporting error at follow-up interviews. J Epidemiol Biostatistics 2001;6(5):393-407.

Schneider A, Oltersdorf T, Schneider V, Gissmann L. Distribution pattern of human papillomavirus 16 genome in cervical neoplasia by molecular in situ hybribization of tissue sections. Int J Cancer 1987; 39:717-21.

Sedjo R, Inserra P, Abrahamsen M, Hariis R, Roe D, Baldwin S, Giuliano A. Human papillomavirus persistence and nutirents involved in the methylation pathway among a cohort of young women. Cancer Epidemiol Biomarkers & Prev 2002a; 11:353-9.

Sedjo R, Roe D, Abrahamsen M, Harris R, Craft N, Baldwin S, Giulinao A. Vitamin A, caretenoids and risk of persistent oncogenic human papillomavirus infection. Cancer Epidemiol Biomarkers & Prev 2002b; 11:876-84.

Sedlacek T, Lindheim S, Eder C, Hasty L, Woodland M , Ludomirsky A, Rando R. Mechanism for human papillomavirus transmission at birth. Am J Obstet Gynecol. 1989; 161:55-9.

Sellors J, Karwalajtys T, Kaczorowski J, Mahony J, Lytwyn A, Chong S, Sparrow J, Lorincz A, for the survey of HPV in Ontario women (SHOW) group . Incidence, clearance and predictors of human papillomavirus infection in women. CMAJ 2003; 168(4):421-5.

Sellors J, Mahony J, Kaczorowski J, Lytwyn A, Bangura H, Chong S, Lorincz A, Dalby D, Janjusevic V, Keller J, for the Survey of HPV in Ontario Women (SHOW) Group. Prevalence and predictors of human papillomavirus infection in women in Ontario, Canada. CMAJ 2000; 163(5):503-8.

Sethi S, Muller M, Schneider A, Blettner M, Smith E , Turek L, Wahrendorf J, Gissman L, Chang-Claude J. Serologic response to the E4, E6, and E7 proteins of human papillomavirus type 16 in pregnant women. Am J Obstet Gynecol 1998; 178:360-64.

Sjalander A, Birgander R, Kivela A, Beckman G. p53 polymorphisms and haplotypes in different ethnic groups. Hum Hered 1995; 45 (3):144-9.

Skrabanek P. Cervical cancer in nuns and prostitutes: a plea for scientific continence. J Clin Epidemiol 1988; 41:577-82.

Smith J, Munoz N, Herrero R, Eluf-Neto J, Ngelangel C, Franceschi S, Bosch F, Walboomers J, Peeling RW. Evidence for Chlamydia trachomatis as an HPV cofactor in the etiology of invasive cervical cancer in Brazil and the Philipines. J Infect Dis 2002; 185:324-31.

Solomon D. The 1988 Bethesda system for reporting cervical/vaginal cytologic diagnoses. Developed and approved at the National Cancer Institute Workshop, Bethesda, Maryland, USA, December, 12-13, 1988. J Clin Cytol Cytopathol 1989; 33:567-74.

Solomon D, Schiffman M, Tarone R, *for the ALTS Group*. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: Baseline results from a randomized trial. J Natl Cancer Inst 2001; 93(4):293-99.

Statistics Canada. 2003; Available at http://www.statcan.ca/english/concepts/nphs/.

Storey A, Thomas M, Kalita A, Harwood C, Gardiol D, Mantovani F, Breuer J, Leigh I, Matlashewski G, Banks L. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. Nature 1998; 393:229-34.

Sun XW, Ellerbrock TV, Lungu O, Chiasson MA, Bush TJ, Wright TCJ. Human papillomavirus infection in human immunodeficiency virus- seropositive women. Obstet Gynecol 1995; 85(5 Pt 1):680-6.

Szarewski A, Cuzick J. Smoking and cervial neoplasia: a review of the evidence. J Epidemiol and Biostatistics 1998; 3:229-56.

The ALTS Group. Human papillomavirus testing for triage of women with cytologic evidence of low-grade squamous intraepithelial lesions: baseline data from a randomized trial. J Natl Cancer Inst 2000; 92(5):397-402.

Thiry L, Vokaer R, Detremmerie O, De Schepper N, Herzog A, Bollen I. Cancer du col uterin, virus du papillome, contraception et tabac. J Gynecol Obstet Biol Reprod 1993; 22:477-86.

Thomas D, Ray R, Kuypers J, Kiviat N, Koetsawang A, Ashley R, Qin Q, Koetsawang S. Human papillomaviruses and cervical cancer in Bangkok. III. The role of husbands and commercial sex workers. Am J Epidemiol 2001; 153 (8):740-48.

Thomas K, Hughes J, Kuypers J, Kiviat N, Lee S, Adam D, Koutsky L. Concurrent and sequential acquisition of different genital human papillomavirus types. J Infect Dis 2000; 182(4):1097-102.

Tjiong M, Out T, Ter Schegget J, Burger M, Van der Vange N. Epidemiologic and mucosal immunologic aspects of HPV infection and HPV related cervial neoplasia in the lower female genital tract: a review. Int J Gynecol Cancer 2001; 11:9-17.

Unger E, Duarte-Franco E. Human papillomaviruses into the new millenium. Obstet Gynecol Clin North America 2001; 28(4):653-66.

van den Brule A, Walboomers J, Dumaine M, Kenemans P, Meijer C. Difference in prevalence of human papillomavirus genotypes in cytomorphologically normal cervical smears is associated with a history of cervical intraepithelial neoplasia. Int J Cancer 1991; 48:404-8.

van Doorn L, Quint W, Kleter B, Molijn A, Colau B, Martin M, Torrez-Martinez K, Peyton C, Wheeler C. Genotyping of Human Papillomavirus in Liquid Cytology Cervical Specimens by the PGMY Line Blot Assay and the SPF_{10} Line Probe Assay . J. Clin. Microbiol. 2002; 40 :979-83.

Van Doornum G, Prins M, Juffermans L, Hooykaas C, van den Hoek J, Coutinho R, Quint W. Regional distribution and incidence of human papillomavirus infections among heterosexual men and women with multiple sexual partners: a prospective study. Genitourin Med 1994; 70:240-6.

van Duin M, Snijders P, Vossen M, Klaassen E, Voorhorst F, Verheijen R, Helmerhorst T, Meijer C, Walboomers J. Analysis of human papillomavirus type 16 E6 variants in relation to p53 codon 72 polymorphism genotypes in cervical carcinogenesis. J Gen Virol 2000; 81:317-25.

van Staveren W, West C, Hoffmans M, Bos P, Kardinaal A, van Poppel G, Schipper H, Hautvast J, Hayes R. Comparison of contemporaneous and retrospective estimates of food consumption made by a dietary history method. Am J Epidemiol 1986; 123:884-93.

Vermund S, Kelly K, Klein R, Feingold A, Schreiber K, Munk G, Burk R. High risk of human papillomavirus infection and cervical squamous intraepithelial lesions among women with symptomatic human immunodeficiency virus infection. Am J Obstet Gynecol 1991; 165:392-400.

Villa L, Franco E. Epidemiologic correlates of cervical neoplasia and risk of human papillomavirus infection in asymptomatic women in Brazil. J Natl Cancer Inst 1989; 81:332-40.

Villa L, Sichero L, Rahal P, Caballero O, Ferenczy A, Rohan T, Franco E. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. J General Virol 2000; 81:2959-68.

von Knebel-Doeberitz M, Oltersdorf T, Schwarz E, Gissman L. Correlation of modified human papillomavirus early gene expression with altered growth properties in C4-1 cervical carcinoma cells. Cancer Res 1988; 48:3780-86.

Vousden K. Interactions of human papillomavirus transforming proteins with the products of tumor suppressor genes. FASEB Journal 1993; 7:872-9.

Walboomers J, de Roda Husman A, van den Brule A, Snijders P, Meijer C. Detection of genital human papillomavirus infections: critical review of methods and prevalence studies in relation to cervical cancer. In: Stern P & S MA . Human papillomaviruses and cervical cancer. Oxford, UK: Oxford University Press, 1994: 41-71.

Walboomers J, Jacobs M, Manos M, Bosch F, Kummer J, Shah K, Snijders P, Peto J, Meijer C, Muńoz N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999; 189:12-9.

Wallin K, Wiklund F, Angstrom T, Bergman F, Stendahl U, Wadell G, Hallmans G, Dillner J. Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. JAMA 1999; 341(22):1633-8.

Wank R, Schendel D, Thomssen C. HLA antigens and cervical cancer. Nature 1992; 356:22-3.

Weiderpass E, Ye W, Tamimi R, Trichopolous D, Nyren O, Vainio H, Adami H-O. Alcoholism and risk of cancer of the cervix uteri, vagina and vulva. Cancer Epidemiol, Biomarkers Prev 2001; 10:899-901.

Werness B, Levine A, Howley P. Association of human papillomavirus type 16 and 18 E6 proteins with P53. Science 1990; 248:76-9.

Wheeler C, Parmenter C, Hunt W, Becker T, Greer C, Hildesheim A, Manos M. Determinants of genital human papillomavirus infection among cytologically normal women attending the University-of-New-Mexico student health center. Sex Transm Dis 1993; 20:286-9.

Wheeler C, Yamada T, Hildesheim A, Jenison S. Human papillomavirus type 16 sequence variants: identification by E6 and L1 lineage-specific hybridization. J Clin Microbiol 1997; 35(1):11-9.

White W, Wilson S, Bonnez W, Rose R, Koenig S, Suzich J. In vitro infection and type-restricted antibody-mediated neutralization of authentic human papillomavirus type 16. J Virol. 1998; 72:959-64.

Winer R, Lee S-K, Hughes J, Adam D, Kiviat N, Koutsky L. Genital human papillomavirus infection: incidence and risk factors in a cohort of female university students. Am J Epidemiol 2003; 157(3):218-26.

Woodman C, Collins S, Winter H, Bailey A, Ellis J, Prior P, Yates M, Rollason T, Young L. Natural history of cervical human papillomavirus infection in young women: a longitudinal cohort study. Lancet 2001; 357:1831-6.

Worley, K, McLeod, M. Clustal Multiple Sequence Alignment 1.8 program. Available at: http://www.geocities.com/ecs124/homework/7/MultipleAlignment.html

Wynder E, Cornfield J, Shroff P, Doraiswami K. A study of environmental factors in carcinoma of the cervix. Am J Obstet Gynecol 1954; 68(4):1016-52.

Xi L, Carter J, Galloway D, Kuypers J, Hughes J , Lee S, Adam D, Kiviat N, Koutsky L. Acquisition and natural history of human papillomavirus type 16 variant infection among a cohort of female university students. Cancer Epidemiol, Biomarkers & Prev 2002; 11 :343-51.

Xi L, Demers W, Koutsky L, Kiviat N, Kuypers J , Watts D, Holmes K, Galloway D. Analysis of human papillomavirus type 16 variants indicates establishment of persistent infection. J Infect Dis 1995; 172:747-55.

Xi L, Koutsky L, Galloway D, Kuypers J, Hughes J , Wheeler C, Holmes K, Kiviat N. Genomic variation of human papillomavirus type 16 and risk for high-grade cervical intraepithelial neoplasia. J Natl Cancer Inst 1997; 89:796-802.

Yamada T, Manos M, Peto J, Greer C, Munoz N, Bosch F, Wheeler C. Human papillomavirus type 16 sequence variation in cervical cancers - a worlwide perspective. J Virol 1997; 71:2463-72.

Yamada T, Wheeler C, Halpern A, Stewart A, Hildesheim A, Jenison S. Human papillomavirus type 16 variant lineages in United Sates populations characterized by nucleotide sequence analysis of the E6, l2 and L1 coding segments. J Virol 1995; 69:7743-53.

Ylitalo N, Josefsson A, Melbye M, Sörenson P, Frisch M, Anderson P, Sparén P, Gustafsson M, Magnusson P, Pontén J, Gyllensten U, Adami H. A prospective study showing long-term infection with human papillomavirus 16 before the development of cervical carcinoma *in situ*. Cancer Res 2000; 60:6027-32.

Zehbe I, Voglino G, Delius H, Wilander E, Tommasino M. Risk of cervical cancer and geographical variations of human papillomavirus 16 E6 polymorphisms. Lancet 1998; 352:1441-2.

Zehbe I, Voglino G, Wilander E, Genta F, Tommasino M. Codon 72 polymorphism of p53 and its association with cervical cancer. Lancet 1999; 354:218-19.

Zielinski G, Snijders P, Rozendaal L, Voorhorst F, van der Linden H, Runsink A, de Schipper F, Meijer C. HPV presence precedes abnormal cytology in women developing cervical cancer and signals false negative smears. Br J Cancer 2001; 85(3):398-404.

zur Hausen H. Condyloma acuminata and human genital cancer. Cancer Res 1976;36:794

zur Hausen H. Papillomaviruses as carcinomaviruses. Adv Viral Oncol 1989; 8:1-36.

zur Hausen H. Papillomavirus causing cancer: evasion from host-cell control in early events in carcinogenesis. J Natl Cancer Inst 2000; 92(9):690-8.

zur Hausen H, Devilliers E. Human papillomaviruses. Annual Review of Microbiology 1994; 48:427-47.

APPENDIX I: MATERIALS & METHODS

APPENDIX I.I: CONSENT FORM

STUDY ON PERSISTENT HPV INFECTION AND CERVICAL INTRAEPITHELIAL NEOPLASIA

Consent Form

I_____ freely consent to participate in the research project of the above title. The following aspects of the study have been explained to me.

A. Purpose of this study

This project conducted by McGill scientists seeks to investigate the occurrence of transient and persistent Human Papillomavirus (HPV) infection and cervical lesions detected by the Pap test in female university students and the related life style risk factors. HPV is a virus (the human papilloma virus) which is normally associated with asymptomatic infections of the genital area. HPV infection is detected by collecting samples of cells from the cervix of the uterus. The sample is then examined to determine the presence of HPV. If HPV is detected, further analysis is conducted to classify the type of HPV.

B. Procedure

If I agree to participate, I will be asked to complete eight self-administered questionnaires, 2 interviews and undergo five Pap smears. The pap specimens will be sent to a lab to be tested for evidence of HPV infection or any alterations that may much later in my life develop into cancer.

C. Personal Inconvenience accompanying participation in this study

It is recommended by Canadian health officials that all sexually active women undergo a Pap smear at least once a year to detect gynecological malignancy. For purposes of this study I will be required to have 5 Pap Smears over a period of 24 months at 6 month intervals. For example,

Jan 97	June 97		Dec. 97	June 98
Dec.98				
1st Pap	2nd Pap		3rd Pap	4th Pap
5th Pap				
1st mth	6 mth	12 mth	18 mth	24 mth

The interviews and self-report questionnaires will be scheduled to coincide with my final two visits. A member of the research team will contact me by phone prior to each scheduled visit to remind me of my appointment and obtain the necessary information for each visit.

D. Compliance

It is imperative for statistical purposes to obtain the lab results of all five Pap Smears as well as responses to the questionnaires. The researchers are aware of the inconveniences of returning to the clinic every six months and will pay you $20 per follow-up return visits. If you complete the entire study you will receive a total of $80.00

E. Risks and Benefits

The risks in this study are minimal as the Pap smear is a safe examination. As with any gynecological examination, there is a possibility that a slight discomfort might be felt during the insertion of the cervical sampler to collect the Pap smear. The benefits of this project include improved characterization of the suspected virus and increase our knowledge of a potentially hazardous disease. As well if any lesions are detected we will notify your doctor at the clinic so that you can be treated if necessary.

F. Confidentiality

In order to ensure my privacy and confidentiality my name will not appear on any record or results. Instead the patient identification number will be assigned to me and will appear on all my records. My patient number will be kept on file at the McGill University Students Health Services clinic and only the investigator and the assistants will have access to the study number. I understand that all information about me or my Pap smear results will be treated in the same confidential manner as other medical records and I will not be identified in any subsequent reporting of results.

G. Voluntary consent

I understand that my participation is entirely voluntary and that I may withdraw at any time without affecting my status at the McGill University Student Health Services Clinic.
By signing this consent form I acknowledge that this research study has been thoroughly explained to me and I fully understand that I will have to commit to making four additional visits once every six months for 24 months. I have the opportunity to ask questions and to seek further information about the procedure and the results of the study. I understand that I am free to ask additional questions in the future and that my identity will remain confidential.

Participant (signature)	Clinical Nurse coordinator (signature)
Print Name	Print Name
Date	Date

APPENDIX I.II: PERSONAL DATA SHEET

PERSONAL DATA SHEET

Please fill in this sheet and sign the attached consent form.
Thank you.

The information contained on this sheet will not belinked to the questionnaire information which will remain anonymous. This information is necessary so that we can contact you for a follow-up appointment at the health service in the future.

Name: _____

Date of Birth: _____
day/month/year

McGill student ID number: _____

Address in Montreal: _____

_____

Telephone Number(s) in Montreal: _____

Permanent Address: _____

_____

Permanent Telephone Number: _____

Do you expect to be studying at McGill during the next 2 years? ***(Circle one)***

Yes **No**

In which Department/Faculty are you enrolled? _____

APPENDIX I.III: BASELINE QUESTIONNAIRE

STUDY NO: _____

WOMEN'S HEALTH STUDY
INITIAL QUESTIONNAIRE

McGill University Student Health Services
Departments of Oncology and Epidemiology & Biostatistics

INSTRUCTIONS FOR THE QUESTIONNAIRE

This questionnaire is composed of the following sections:

General information
Diet History
Smoking history and alcohol consumption
Reproductive history
Sexual history
Contraceptive history
Personal hygiene habits
Medical history

Most questions require that you simply check a box ❑ with an "X" to indicate your choice. Other questions require a specific answer, such as age, date, or another number. Depending on your answer for some questions, you will be told to skip the next question and go to a different part of the questionnaire. This is to save you time, so that you won't have to go over questions that do not apply to you.

There are no right or wrong answers to any question. Many questions require that you think back over your adult years, particularly over the past year, to recall specific information. Please take the time to reflect. If you prefer, you can answer sections of the questionnaire on different days. If you choose to do so, check your answers from previous days to make sure you agree with them before mailing the questionnaire back to us. You will be surprised that by being "forced" to recall specific information of one type, some of the answers for other questions may come more naturally to you later on. If you can't possibly remember the information skip the question, but we would like to encourage you to try to answer all questions. A good guess is always better than no information at all. If you'd like to tell us more about any specific items please use the available space at the end of the questionnaire.

WE APPRECIATE YOUR COOPERATION WITH THE STUDY

STUDY NO: _____

GENERAL INFORMATION

This portion of the questionnaire concerns general information about you and where you live.

1. What is your date of birth? _____/_____/_____ ***(very important)***
D M Y

2. In what country were you born? _____

↳ If born in Canada: indicate province: _____

3. What is your **current** marital status?

- ❑ Married
- ❑ Single
- ❑ Unmarried, but living with a partner
- ❑ Divorced/separated
- ❑ Widowed

4. The Montreal area is made up of many ethnic groups. We would like to know in which group you would place yourself. Check the most appropriate category:

- ❑ French Canadian
- ❑ English Canadian
- ❑ Black Canadian
- ❑ Native Indian
- ❑ Jewish
- ❑ Hispanic/Portuguese
- ❑ Greek
- ❑ Italian
- ❑ Asian/Oriental
- Other: _____

5. a) What is/was your father's occupation? _____

b) What is/was your mother's occupation? _____

c) Would you say that your family's financial situation while growing up was:

❑ Difficult ❑ Moderate ❑ Very comfortable

6. How are you presently enrolled at McGill?

- ❑ Undergraduate - Regular Student (State year: _____)
- ❑ Graduate studies - Diploma, Master's or Doctoral Program
- ❑ Other (e.g. Trainee, Postdoctoral Studies, Sabbatical)

DIET HISTORY

This section of the questionnaire concerns some specific food and beverage items. We want to know about your usual adult diet, that is, your usual eating habits ***during all your adult years.***

7. For each item check the category that best reflects your average consumption pattern. Try a good guess considering a typical serving and any cooking method:

	At least once a day	<1 per day >1 per week	At least once a week	At least once a month	Never
a) carrots	❑	❑	❑	❑	❑
b) spinach	❑	❑	❑	❑	❑
c) broccoli	❑	❑	❑	❑	❑
d) lettuce	❑	❑	❑	❑	❑
e) cabbage	❑	❑	❑	❑	❑
f) cheese or cream	❑	❑	❑	❑	❑
g) milk or yoghurt	❑	❑	❑	❑	❑
h) liver	❑	❑	❑	❑	❑
I) pure orange juice	❑	❑	❑	❑	❑
j) fresh orange or grapefruit	❑	❑	❑	❑	❑
k) vitamin C-fortified fruit drink such as tropical fruit juice, or Tang	❑	❑	❑	❑	❑
l) vitamin-C supplements	❑	❑	❑	❑	❑
m) mixed vegetable juices (V-8, garden cocktail, tomato juice)	❑	❑	❑	❑	❑

SMOKING HISTORY AND ALCOHOL CONSUMPTION

The following questions are about your tobacco smoking and alcoholic beverage consumption habits. Please try to be as specific as possible in your answers.

8. Have you ever smoked cigarettes **regularly**, that is, **one** cigarette or **more** each **day** for a **year** or more?

❑ Yes ❑ No

9. Have you smoked a total of at least 100 cigarettes in your lifetime?

❑ Yes ❑ No → **If No**, go to question 13

10. At what age did you start to smoke? _____years

11. Do you still smoke? ❑ Yes ❑ No

↓

If No, at what age did you stop?____

12. On average, how many cigarettes do/did you smoke a day?_____

13. Has there ever been a period in your life when you drank beer, wine or liquor
AT LEAST ONCE A MONTH?

❑ Yes ❑ No

↳ **If No**, go to question 15

14. Has there ever been a period in your life when you drank beer, wine or liquor
AT LEAST ONCE A WEEK?

❑ Yes ❑ No

↓ ↳ **If No**, go to question 15

If Yes,
indicate the average number of drinks per week that you consumed during the past 5 years *(consider a drink as being equivalent to a 12 oz. can of beer or to a 4 oz. glass of wine or to 1.5 ounces of hard liquor such as gin, vodka, whiskey, scotch, rum, tequilla, etc.).*

a) Beer: _____ cans per week

b) Wine: _____ glasses per week

c) Liquor: _____ drinks per week

REPRODUCTIVE HISTORY

In this section of the questionnaire we would like to know about your reproductive health including all your pregnancies as well as miscarriages and abortions.

15. At what age did you have your first menstrual period?_____years

16. To the best of your knowledge, are you currently pregnant?

❑ Yes ❑ No ❑ Don't know

17. Have you ever been pregnant before?

❑ Yes ❑ No

↓ → **If No**, go to question 20

If Yes,
how many times?_____ times

18. How many of your pregnancies resulted in:

a) livebirths: _____

b) stillbirth: _____

c) miscarriage: _____

d) abortion: _____

19. How many of your full-term pregnancies resulted in:

a) vaginal deliveries: _____

b) cesarean sections: _____

SEXUAL HISTORY

The next questions are about your sexual history. We realize this is a personal subject, but it is very important to the study. Please take the time to recall this information as accurately as possible. Note that some questions in this section refer to your entire life as an adult, whereas others refer only to your recent experience. We would like to remind you that all the information you give us will be kept entirely confidential.

20. Have you ever engaged in vaginal sexual intercourse?

❑ Yes ❑ No

↓ └→ **If No**, go to question 27

If Yes,
how old were you when you first had vaginal sexual intercourse?___

21. THROUGHOUT YOUR LIFE, what is the number of male partners with whom you have had vaginal sexual intercourse?

Number *(approximately)*_____

22. With how many of these male partners did you have a sexual relationship involving intercourse on a regular basis for **three months or longer?**

Number_____ ❑ None

23. For MOST OF YOUR SEXUALLY ACTIVE LIFE, how often on the **average**, did you have vaginal sexual intercourse? Please give your answer in number of times per week, month, or year, whichever is easiest:

Number of times per week_____
OR
Number of times per month_____
OR
Number of times per year _____
OR
Less than once a year ❑

24. During THE LAST YEAR ONLY, what is the number of male partners with whom you have had vaginal sexual intercourse?

_____ Number ❑ None in the past year

How many of those partners were new?_____ Number

25. In THE LAST YEAR ONLY, how often on the **average**, did you have vaginal sexual intercourse? Please give your answer in number of times per week, month, or year, whichever is easiest:

Number of times per week_____
OR
Number of times per month_____
OR
Number of times per year _____
OR
Less than once a year ❑

26. When you are having your menstrual periods, do you have vaginal sexual intercourse?

❑ Yes ❑ No

27. THROUGHOUT YOUR LIFE, has anyone ever performed oral sex on you?

_____ Number ❑ None

→ **If No**, go to question 31

28. How often on **average**, did you receive oral sex? Please give your answer in number of times per week, month, or total per year, whichever is easiest:

Number of times per week_____
OR
Number of times per month_____
OR
Number of times per year_____
OR
Less than once a year ❑

29. During THE LAST YEAR ONLY, how many people performed oral sex on you?

_____ Number ❑ None in the past year

30. During THE LAST YEAR ONLY, how often on **average**, did you receive oral sex? Please give your answer in number of times per week, month, or year, whichever is easiest:

Number of times per week_____
OR
Number of times per month_____
OR
Number of times per year _____
OR
Less than once a year ❑

31. Do you ever practice anal intercourse?

❑ Yes ❑ No

↳ **If yes**, would you say that you have had anal intercourse:

❑ Frequently ❑ Occasionally ❑ Rarely

32. With whom do you usually have sex?

❑ Men ❑ Women ❑ Both

33. Do you ever masturbate?

❑ Yes ❑ No

↳ **If yes**, do you ever insert objects into the vagina for stimulation:

❑ Yes ❑ No

CONTRACEPTIVE HISTORY

Here we would like to know about methods of birth control or family planning that you and your husband/partner used. It would be important to indicate all the methods you've used since you became sexually active. ***If you answered "No" to question 20 you may skip this section entirely and go to question 38.***

34. The following is a list of common birth control methods. Read along the list and check if you and a sex partner have ever used any of them (check all that apply) either occasionally or regularly.

BY REGULARLY WE MEAN AT LEAST 3 MONTHS CONSECUTIVELY

a) oral contraceptive (birth control pill)	❑ Regularly	❑ Sometimes	❑ Never
b) condom (rubber)	❑ Regularly	❑ Sometimes	❑ Never
c) foam, jelly, cream, or suppository	❑ Regularly	❑ Sometimes	❑ Never
d) loop, coil, or other intrauterine device	❑ Regularly	❑ Sometimes	❑ Never
e) diaphragm	❑ Regularly	❑ Sometimes	❑ Never
f) cervical cap	❑ Regularly	❑ Sometimes	❑ Never
g) sponge	❑ Regularly	❑ Sometimes	❑ Never
h) vaginal douche	❑ Regularly	❑ Sometimes	❑ Never
i) rhythm, calendar, or natural method	❑ Regularly	❑ Sometimes	❑ Never
j) withdrawal/pulling out	❑ Regularly	❑ Sometimes	❑ Never

35. If you have used oral contraceptives or birth control pills, please indicate how old you were when you first took them?

Age:_____ years ❑ Never used oral contraceptives

Go to question 37

36. Considering only the times when you were taking the pill, for how long have you been relying on this method of birth control (add together all periods during which you took any oral contraceptives)?

_____months

OR

_____years

OR

❑ all periods combined were less than 3 months

37. Now, considering ONLY THE LAST YEAR, on the average, which of the following birth control methods have you or your partner come to rely upon? **(check all that apply)**

BY REGULARLY WE MEAN AT LEAST 3 MONTHS CONSECUTIVELY

I was not sexually active ❑ ⟶ Go to Question 38

a) oral contraceptive (birth control pill)	❑ Regularly	❑ Sometimes	❑ Never
b) condom (rubber)	❑ Regularly	❑ Sometimes	❑ Never
c) foam, jelly, cream, or suppository	❑ Regularly	❑ Sometimes	❑ Never
d) loop, coil, or other intrauterine device	❑ Regularly	❑ Sometimes	❑ Never
e) diaphragm	❑ Regularly	❑ Sometimes	❑ Never
f) cervical cap	❑ Regularly	❑ Sometimes	❑ Never
g) sponge	❑ Regularly	❑ Sometimes	❑ Never
h) vaginal douche	❑ Regularly	❑ Sometimes	❑ Never
i) rhythm, calendar, or natural method	❑ Regularly	❑ Sometimes	❑ Never
j) withdrawal/pulling out	❑ Regularly	❑ Sometimes	❑ Never

PERSONAL HYGIENE HABITS

This section concerns your personal hygiene habits. As with other topics in this questionnaire, this is also a personal subject of great importance to the study. Again, rest assured that we will treat your answers confidentially. Please take the time to recall this information as accurately as possible. A good guess is always better than leaving the question blank.

38. How many times per day or per week do you usually bathe or shower?

_____per day
OR
_____per week

39. **Aside from** those baths and showers, do you ever wash your genital area? (Do not consider the times you may wash after sexual intercourse.)

❑ Yes ❑ No

↓

If Yes, in the **last year**, on the average, how many times per day, week, or month did you wash your genital area?

_____per day
OR
_____per week
OR
_____per month
OR
❑ less than once a month

40. Have you ever used a vaginal douche?

❑ Yes ❑ No

↓ → **If No**, go to question 42

If Yes,
in the **last year**, on the average, how many times per day, week, or month did you use a vaginal douche?

_____per day
OR
_____per week
OR
_____per month
OR
❑ less than once a month

41. **If you answered yes in question 40**, what kind of douche did you use?

❑ Water only ❑ Water and vinegar ❑ Other (*specify:*_____)

42. Have you ever used a feminine genital spray or vaginal deodorant?

❑ Yes ❑ No

↓

If Yes,in the **last year**, on the average, how many times per day, week, or month did you use a vaginal deodorant?

_____per day
OR
_____per week
OR
_____per month
OR
❑ less than once a month

43. When you are menstruating, what do you use to collect the blood? (check all that apply)

❑ sanitary pads/napkins
❑ tampons
❑ other

44. Following sexual intercourse, do you usually wash your genital area within the hour? Choose the category that best reflects your behaviour during most of your adult life: (skip this question if you never had sexual intercourse).

❑ always
❑ sometimes
❑ rarely
❑ never

45. Following oral sex, do you usually wash your genital area within the hour? Choose the category that best reflects your behaviour during most of your adult life: (skip this question if you never practice oral sex).

❑ always
❑ sometimes
❑ rarely
❑ never

46. What is your preferred hygiene practice after each bowel movement? (Check all that apply)

❑ Use toilette paper with a BACK-TO-FRONT hand motion for wiping
❑ Use toilette paper with a FRONT-TO-BACK hand motion for wiping
❑ Wash with water only
❑ Wash with water and soap

MEDICAL HISTORY

The next questions are about the frequency with which you have taken PAP smears and about some medical problems including sexually transmitted diseases. We realize that this is a sensitive subject but, again, it is very important to the research. We appreciate your honesty and want to remind you that all information you give us is kept private and confidential.

47. Thinking back over your adult years, how often have you usually had a PAP smear? Choose one category below:

- ❑ this is my first PAP smear
- ❑ 2-3 times
- ❑ 4-5 times
- ❑ 6-10 times
- ❑ more than 10 times

48. What is the month and year of the last PAP smear you had? _____/_____
Month Year

49. Did a doctor ever tell you that you had one of the following conditions? Check all that apply, if you are in doubt check the "don't know" column.

a) Vaginal yeast infections:	❑ Yes	❑ No	❑ Don't know
b) Trichomonas vaginal infections:	❑ Yes	❑ No	❑ Don't know
c) Venereal warts, condylomas, or papilloma virus infections:	❑ Yes	❑ No	❑ Don't know
d) Chlamydia:	❑ Yes	❑ No	❑ Don't know
e) Genital herpes:	❑ Yes	❑ No	❑ Don't know
f) Syphilis:	❑ Yes	❑ No	❑ Don't know
g) Gonorrhea:	❑ Yes	❑ No	❑ Don't know
h) Ulcers or genital sores:	❑ Yes	❑ No	❑ Don't know

50. Thinking back over **all your adult life**, have you experienced other genital conditions such as vaginal discharge, itching or irritation?

- ❑ Never
- ❑ Less than once a year
- ❑ More than once a year

51. Now, **only during the last year**, have you experienced other genital conditions such as vaginal discharge, itching or irritation?

- ❑ Never
- ❑ Once or twice
- ❑ More than 3 times last year

52. Sometimes women are given female hormones by their doctors because of a variety of reasons (alleviate acne, regulate or eliminate painful periods, menopausal symptoms, reduce discomfort during intercourse due to vaginal dryness, prevent miscarriage, among others). To the best of your recollection, were you ever prescribed any female hormones by your doctor?

❑ Yes ❑ No

↓ → **If No**, go to question 55

If Yes,
in what month and year did you start taking them and also, in what month and year did you last take them?

Start: _____/_____ End: _____/_____
month year month year

53. Between the above two dates, for how long (number of months) did you take the female hormone medication on a continual basis, altogether?

_____months

54. Was the female hormone medication in the form of (check all that apply):

- ❑ pills
- ❑ shots
- ❑ creams or suppositories

55. Would you please indicate the date when you finished filling in the questionnaire?

_____/_____/_____
DAY MONTH YEAR

USE THE SPACE BELOW IF YOU HAVE ANY ADDITIONAL INFORMATION YOU FEEL WOULD BE IMPORTANT FOR US TO KNOW:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

This is the end of the questionnaire. We would like you to take a few seconds to review your answers in all sections of the questionnaire. If you answered the sections on different days, take a moment to reflect if you agree now with your answers from previous days. Again, try to answer all questions; a good guess will be more useful to the study than leaving the question blank.

THANK YOU VERY MUCH FOR YOUR COOPERATION

APPENDIX I.IV: FOLLOW-UP QUESTIONNAIRE

STUDY NO: _____

WOMEN'S HEALTH STUDY FOLLOW-UP QUESTIONNAIRE

McGill University Student Health Services

Departments of Oncology and Epidemiology & Biostatistics

INSTRUCTIONS FOR THE QUESTIONNAIRE

This questionnaire is composed of the following sections:

General information
Sexual history
Contraceptive history
Personal hygiene habits
Medical history

Most questions require that you simply check a box ❑ with an "X" to indicate your choice. Other questions require a specific answer, such as age, date, or another number. Depending on your answer for some questions, you will be told to skip the next question and go to a different part of the questionnaire. This is to save you time, so that you won't have to go over questions that do not apply to you.

Many questions also refer to the period since your last visit to the clinic, a few months ago, when you were given a similar questionnaire. In those instances the questions will start with "since your last visit... ". There are no right or wrong answers to any question. Many questions require that you think back over your adult years, particularly over the past year, to recall specific information. Please take the time to reflect. If you prefer, you can answer sections of the questionnaire on different days. If you choose to do so, check your answers from previous days to make sure you agree with them before mailing the questionnaire back to us. You will be surprised that by being "forced" to recall specific information of one type, some of the answers for other questions may come more naturally to you later on. If you can't possibly remember the information skip the question, but we would like to encourage you to try to answer all questions. A good guess is always better than no information at all. If you'd like to tell us more about any specific items please use the available space at the end of the questionnaire. Once you have completed the questionnaire please return it to the study nurse at the clinic.

WE APPRECIATE YOUR COOPERATION WITH THE STUDY

STUDY NO: _____

GENERAL INFORMATION

This portion of the questionnaire concerns general information about you and where you live.

1. What is your date of birth? _____/_____/_____ ***(very important)***
D M Y

2. In what country were you born? _____

↳ If born in Canada: indicate province: _____

3. What is your **current** marital status?

- ❑ Married
- ❑ Unmarried, but living with a partner
- ❑ Widowed
- ❑ Single
- ❑ Divorced/separated

4. On average, how many cigarettes have you smoked **since your last visit**?

Number of cigarettes: _____ per day OR _____ per week OR ❑ Non-smoker

5. Indicate the average number of drinks per week that you consumed **since your last visit** *(consider a drink as being equivalent to a 12 oz. can of beer or to a 4 oz. glass of wine or to 1.5 ounces of hard liquor such as gin, vodka, whiskey, scotch, rum, tequilla, etc.).*

a) Beer: _____ cans per week

b) Wine: _____ glasses per week

c) Liquor: _____ drinks per week

d) None since last visit ❑

SEXUAL HISTORY

The next questions are about your sexual history. We realize this is a personal subject, but it is very important to the study. Please take the time to recall this information as accurately as possible. Note that some questions in this section refer to your entire life as an adult, whereas others refer only to your recent experience. We would like to remind you that all the information you give us will be kept entirely confidential.

6. Have you ever engaged in vaginal sexual intercourse?

❑ Yes ❑ No → **If No**, go to question 11

↓

If Yes,

how old were you when you first had vaginal sexual intercourse? _____ years

7. THROUGHOUT YOUR LIFE, what is the number of male partners with whom you have had vaginal sexual intercourse?

Number *(approximately)* _____

8. SINCE YOUR LAST VISIT, what is the number of male partners with whom you have had vaginal sexual intercourse?

_____ Number ❑ None since last visit

↳ How many of those partners were new? _____ Number

9. SINCE YOUR LAST VISIT, how often on the **average**, did you have vaginal sexual intercourse? Please give your answer in number of times per week, or month, whichever is easiest:

Number of times per week _____

OR

Number of times per month _____

OR

Never since last visit ❑

10. When you are having your menstrual periods, do you have vaginal sexual intercourse?

❑ Yes ❑ No

11. SINCE YOUR LAST VISIT, how many people performed oral sex on you?

_____ Number ❑ None since last visit

↳ How many of those partners were new? _____ Number

12. SINCE YOUR LAST VISIT, how often on **average**, did you receive oral sex? Please give your answer in number of times per week, or month, whichever is easiest:

Number of times per week _____
OR
Number of times per month _____
OR
Never since last visit ❑

13. SINCE YOUR LAST VISIT, have you practiced anal intercourse?

❑ Yes ❑ No

↳ **If yes**, would you say that you have had anal intercourse:

❑ Frequently ❑ Occasionally ❑ Rarely

14. SINCE YOUR LAST VISIT, with whom do you usually have sex?

❑ Men ❑ Women ❑ Both

15. Do you ever masturbate?

❑ Yes ❑ No

↳ **If yes**, do you ever insert objects into the vagina for stimulation:

❑ Yes ❑ No

16. To the best of your knowledge, are you currently pregnant?

❑ Yes ❑ No ❑ I don't know

CONTRACEPTIVE HISTORY

Here we would like to know about methods of birth control or family planning that you and your husband/partner used. ***If you answered "No" to question 6 you may skip this section entirely and go to question 18.***

17. SINCE YOUR LAST VISIT, on the average, which of the following birth control methods have you and your partner come to rely upon? **(check all that apply)**

BY REGULARLY WE MEAN AT LEAST 3 MONTHS CONSECUTIVELY

I was not sexually active ❑ ⟶ Go to Question 18

a) oral contraceptive (birth control pill)	❑ Regularly	❑ Sometimes	❑ Never
b) condom (rubber)	❑ Regularly	❑ Sometimes	❑ Never
c) foam, jelly, cream, or suppository	❑ Regularly	❑ Sometimes	❑ Never
d) loop, coil, or other intrauterine device	❑ Regularly	❑ Sometimes	❑ Never
e) diaphragm	❑ Regularly	❑ Sometimes	❑ Never
f) cervical cap	❑ Regularly	❑ Sometimes	❑ Never
g) sponge	❑ Regularly	❑ Sometimes	❑ Never
h) vaginal douche	❑ Regularly	❑ Sometimes	❑ Never
i) rhythm, calendar, or natural method	❑ Regularly	❑ Sometimes	❑ Never
j) withdrawal/pulling out	❑ Regularly	❑ Sometimes	❑ Never

PERSONAL HYGIENE HABITS

This section concerns your personal hygiene habits. As with other topics in this questionnaire, this is also a personal subject of great importance to the study. Again, rest assured that we will treat your answers confidentially. Please take the time to recall this information as accurately as possible. A good guess is always better than leaving the question blank.

18. Aside from baths and showers, do you ever wash your genital area? (Do not consider the times you may wash after sexual intercourse.)

❑ Yes ❑ No

↓

If Yes, since your last visit, on the average, how many times per day, week, or month did you wash your genital area?

_____ per day
OR
_____ per week
OR
_____ per month
OR

19. SINCE YOUR LAST VISIT, following vaginal sexual intercourse, do you usually wash your genital area within the hour? Choose the category that best reflects your behaviour, since your last visit: (skip this question if you never had sexual intercourse).

❑ always
❑ sometimes
❑ rarely
❑ never

20. SINCE YOUR LAST VISIT, following oral sex, do you usually wash your genital area within the hour? Choose the category that best reflects your behaviour, since your last visit: (skip this question if you never practice oral sex).

❑ always
❑ sometimes
❑ rarely
❑ never

21. SINCE YOUR LAST VISIT, did you use a vaginal douche?

❑ Yes ❑ No

If Yes, how many times? _____

What kind of douche did you use?

❑ Water only ❑ Water and vinegar ❑ Other (*specify:*_____)

MEDICAL HISTORY

The next questions are about some medical problems including sexually transmitted diseases. We realize that this is a sensitive subject but, again, it is very important to the research. We appreciate your honesty and want to remind you that all information you give us is kept private and confidential.

22. SINCE YOUR LAST VISIT, did a doctor tell you that you had one of the following conditions? Check all that apply, if you are in doubt check the "don't know" column.

a)	Vaginal yeast infections:	❑ Yes	❑ No	❑ Don't know
b)	Trichomonas vaginal infections:	❑ Yes	❑ No	❑ Don't know
c)	Venereal warts, condylomas, or papilloma virus infections:	❑ Yes	❑ No	❑ Don't know
d)	Chlamydia:	❑ Yes	❑ No	❑ Don't know
e)	Genital herpes:	❑ Yes	❑ No	❑ Don't know
f)	Syphilis:	❑ Yes	❑ No	❑ Don't know
g)	Gonorrhea:	❑ Yes	❑ No	❑ Don't know
h)	Ulcers or genital sores:	❑ Yes	❑ No	❑ Don't know

23. SINCE YOUR LAST VISIT, have you experienced other genital conditions such as vaginal discharge, itching or irritation?

- ❑ Never since last visit
- ❑ Once or twice
- ❑ More than 3 since last visit

24. Would you please indicate the date when you finished filling in the questionnaire?

_____/_____/_____
DAY MONTH YEAR

USE THE SPACE BELOW IF YOU HAVE ANY ADDITIONAL INFORMATION YOU FEEL WOULD BE IMPORTANT FOR US TO KNOW:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

This is the end of the questionnaire. We would like you to take a few seconds to review your answers in all sections of the questionnaire. If you answered the sections on different days, take a moment to reflect if you agree now with your answers from previous days. Again, try to answer all questions; a good guess will be more useful to the study than leaving the question blank.

THANK YOU VERY MUCH FOR YOUR COOPERATION

APPENDIX I.V: LIST OF STUDY VARIABLES

Table I.i
Description of the original and recoded variables of interest

Variable [frequency of missing values before variable was recoded]	Classification & distribution at baseline (frequency of exposure)
Site of recruitment (clinic)	McGill (421) Concordia (200)
Age ***Baseline*** **[0]**	17-20 (187) 21-23 (228) 24-26 (102) 27+ (104) *range: 17-45; Mean= 23.0, σ=4.0*
Race ***Baseline*** **[13]**	
Grouped according to following classifications: White (*French/English Canadian, Jewish Canadian, American, European, Australian, New Zealand*) Asian (*Asian/Oriental*), Black (*Black Canadian, African American, Caribbean, African*), Hispanic (*Central & South American, Spanish*)	White (513) Asian (62) Black (29) Hispanic (17)
Diet ***Baseline***	
Average dairy consumption (composite of *milk* &/or *yoghurt* &/or *cheese*) [2]	1+/week in adult lifetime (600) Never in adult lifetime (21)
Average vegetable consumption (composite of *carrots* &/or *spinach* &/or *broccoli* &/or *lettuce* &/or *cabbage*) [2]	1+/day in adult lifetime (132) <1/day - >1/week in adult lifetime (489)
Smoking history and alcohol consumption ***Time-dependent***	
Smoking status (computed from the following variables: *Ever smoke, Still smoking, Age stopped smoking*). *Updated at each subsequent visit, based, in part, on value at previous visit* [0]	Never in lifetime (374) Former in lifetime (98) Current in lifetime (149)
*Number cigarettes smoked on average per day, since last visit.	0-<1/day since last visit (397) 1-5/day since last visit (99) >5/day since last visit (125) *range: 0-30; Mean= 2.9, σ=5.3 (all)* *range: 0.5-30; Mean= 7.2,* *σ=6.2(smokers only)*

Table I.i (continued)

Variable [frequency of missing values before variable was recoded]	**Classification & distribution at baseline (frequency of exposure)**
Lifetime cumulative tobacco exposure was estimated with pack-years of cigarettes smoked. A pack-year was defined as the cumulative exposure equivalent to smoking one pack of cigarettes (20 cigarettes) daily during one year. *Updated at each subsequent visit, based, in part, on value at previous visit*	0 pack-years in lifetime (378) <1 pack-year in lifetime (101) 1-2 pack-years in lifetime (58) >2 pack-years in lifetime (84) *range: 0-22; Mean= 1.0, σ=2.5 (all)* *range: 0-22; Mean= 2.6, σ=3.5 (smokers only)*
*Weekly alcohol consumption since last visit (6 months). (Alcoholic beverage=1 beer or 1 glass wine or 1 ounce of hard liquor.) [3]	0 drinks/week since last visit (232) 1-3 drinks/ week since last visit (173) >3 drinks/ week since last visit (216) *range: 0-63; Mean= 3.3, σ=5.0 (all)* *range: 0.5-63.0; Mean= 5.2, σ=5.5 (drinkers only)*
Sexual History: *Time-independent*	
Age at first intercourse [53]	19+ years (174) 16-18 years (314) <16 years (133) *range: 12-26; Mean= 17.3, σ=2.2*
Number of lifetime partners (baseline value) [12]	1 lifetime partner (147) 2-4 lifetime partners (196) 5-9 lifetime partners (161) 10+ lifetime partners (117) range: 1-75; Mean= 6.0, σ=7.4
Time-dependent	
*New recent number of partners since last visit (6 months). [16]	0 partners since last visit (260) 1 partner since last visit (202) 2+ partners since last visit (159) *range: 0-15; Mean= 1.1, σ=1.4*
*Recent frequency of vaginal sex since last visit (6 months). [18]	<1/week since last visit (111) 1-2/week since last visit (266) 3+/week since last visit (244) *range: 0-13.8; Mean= 2.3, σ=1.9*
Contraceptive history: *Time-independent*	
Time on oral contraceptives (computed from the following variables: *Ever use OC's, Age started using the pill, Age stopped using the pill, Total months/yrs on the pill and if still on pill at subsequent visit, number of months between visits added to value from previous visit) Updated at each subsequent visit, based, in part, on value at previous visit* [26]	Never used OCs in lifetime (145) <1 year in lifetime (121) 1-5 year in lifetime (252) 5+ year in lifetime (103) *range: 1-17; Mean= 4.2, σ=3.0*

Table I.i (continued)

Variable [frequency of missing values before variable was recoded]	Classification & distribution at baseline (frequency of exposure)
Time-dependent	
*Recent oral contraceptive use since last visit (6 months). [26]	Never since last visit (145) Sometimes since last visit (64) Always since last visit (412)
*Recent condom use since last visit (6 months). [21]	Never since last visit (54) Sometimes since last visit (208) Always since last visit (359)
*Recent foam (lubrication or jelly) use since last visit (6 months). [67]	Never since last visit (509) Sometimes since last visit (84) Always since last visit (28)
Personal hygiene habits ***Time-independent***	
Menstrual products usually used in lifetime. [8]	Sanitary pads (107) Tampons (135) Pads & tampons (379)
*Washing within I hour after vaginal or oral sex, since last visit. *Never* and *rarely* categories combined. [12]	Never/rarely since last visit (208) Sometimes since last visit (361) Always since last visit (52)
Medical history ***Time-independent***	
Lifetime number of Pap smears [4]	1 Pap smear in life (109) 2-3 Pap smears in life (224) 4-5 Pap smears in life (115) 6-10 Pap smears in life (99) >10 Pap smears in life (74)
Time-dependent	
**Recent history of Chlamydia infection, reported by a physician since last visit (6 months). [22]	No since last visit (589) Yes since last visit (32)
**Recent history of herpes infection, reported by a physician since last visit (6 months). [24]	No since last visit (603) Yes since last visit (18)
**Recent history of warts infection, reported by a physician since last visit (6 months). [20]	No since last visit (537) Yes since last visit (84)
***Recent frequency of vaginal irritation since last visit (6 months). [6]	Never since last visit (101) <3 times since last visit (315) 3+ time since last visit (205)

**Baseline question referred to average number or use in adult lifetime,*
***Baseline question referred to history in adult lifetime,*
****Baseline question referred to frequency per year in adult lifetime*

APPENDIX II: ADDITIONAL ANALYSES FOR MANUSCRIPT I

Table II.i
Prevalence and incidence of infection for all HPV types tested in the cohort, and for HPV groups according to oncogenicity

HPV type	Baseline Prevalence (%)	Number of incident cases	Women months of follow-up	Incidence rate (per 1000 woman-months) (95% CI)
6	2.7	29	12709	2.3 (1.5, 3.3)
11	0.9	4	13613	0.3 (0.0, 0.6)
16	7.0	62	11928	5.2 (4.0, 6.7)
18	3.1	24	12735	1.9 (1.2, 2.8)
26	0.3	3	13693	0.2 (0.0, 0.45)
31	2.6	21	12854	1.6 (1.0, 2.5)
33	1.1	12	13513	0.9 (0.4, 1.4)
35	0.2	1	13722	0.1 (0.0, 0.2)
39	1.0	24	13476	1.8 (1.1, 2.5)
40	0.0	2	13755	0.2 (0.0, 0.40)
42	0.3	5	13687	0.4 (0.1, 0.7)
45	2.0	7	13051	0.5 (0.2, 1.1)
51	2.9	43	12588	3.4 (2.5, 4.6)
52	2.9	18	12757	1.4 (0.8, 2.2)
53	4.3	31	12468	2.5 (1.7, 3.5)
54	2.7	32	12783	2.5 (1.7, 3.5)
55	0.5	11	13600	0.8 (0.3, 2.3)
56	2.6	19	12842	1.5 (0.9, 2.3)
57	0	0	0	0
58	2.6	12	12931	0.9 (0.5, 1.6)
59	0.3	12	13629	0.9 (0.4, 1.4)
66	1.2	24	13032	1.8 (1.2, 2.7)
68	0.5	9	13581	0.7 (0.3, 1.1)
73 (MM9)	0.8	17	13486	1.3 (0.7, 1.9)
82 (MM4)	0.8	13	13485	1.0 (0.5, 1.50
83 (MM7)	0.5	10	13573	0.7 (0.2, 1.2)
84 (MM8)	3.8	46	12475	3.7 (2.7, 4.9)
Any HPV	**29.0**	**155**	**8151**	**19.0 (16.1, 22.3)**
HR-HPV	**21.8**	**128**	**9344**	**13.7 (11.4, 16.3)**
LR-HPV	**14.8**	**128**	**10299**	**12.4 (10.4, 14.8)**

Table II.ii

Actuarial analysis* of the time to acquisition of the most common HPV types

	Cumulative rate (%) of infection and respective 95% confidence intervals by time since enrollment		
HPV type	**6 months**	**12 months**	**24 months**
HPV-6	0.2 (0.2, 0.6)	2.0 (0.8, 3.2)	4.0 (2.2, 5.8)
HPV-16	1.5 (0.5, 2.5)	4.0 (2.2, 5.8)	12.0 (9.0, 14.9)
HPV-18	0.5 (0.1, 1.0)	1.6 (0.2, 3.4)	4.3 (2.5, 6.1)
HPV-31	0.2 (0.2, 0.6)	2.0 (0.8, 3.2)	3.7 (1.9, 5.5)
HPV-45	0.2 (0.2, 0.6)	0.5 (0.1, 1.0)	1.1 (0.1, 2.1)
HPV-51	0.7 (0.1, 1.5)	2.9 (1.5, 4.3)	7.6 (5.6, 9.6)
HPV-52	0.5 (0.1, 1.1)	1.6 (0.6, 2.6)	3.5 (1.9, 5.1)
HPV-53	0.5 (0.1, 1.1)	3.2 (1.6, 4.8)	5.3 (3.3, 7.3)
HPV-54	0.7 (0.1, 1.5)	2.3 (0.9, 3.7)	4.4 (3.1, 6.7)
HPV-56	0.5 (0.1, 1.1)	1.8 (0.6, 3.0)	3.2 (1.6, 4.8)
HPV-58	0.3 (0.3, 0.9)	0.9 (0.1, 1.7)	2.0 (0.6, 3.4)
HPV-84	0.5 (0.1, 1.1)	4.3 (2.5, 6.1)	7.6 (5.6, 9.6)
Any HPV	**4.6 (2.6, 6.6)**	**18.0 (14.1, 21.9)**	**36.4 (31.3, 41.5)**
HR-HPV	**4.1 (2.3, 5.9)**	**12.9 (9.8, 16.0)**	**29.2 (24.7, 33.7)**
LR-HPV	**2.2 (0.9, 3.5)**	**13.6 (10.5, 16.7)**	**23.9 (19.8, 28.0)**

* *Estimated via the Kaplan-Meier technique.*

Table II.iii

Average duration of type-specific **incident** HPV infections with 1- and 2-year probabilities of remaining HPV positive

HPV type	Median retention* time (95%CI) in months	Mean retention* time (95%CI) in months	Proportion* emaining positive at: 1 year (95%CI)	Proportion* remaining positive at: 2 years (95%CI)
HPV-6	6.4 (4.9, 7.8)	8.7 (6.8, 10.6)	0.42 (0.19, 0.65)	0.16 (0.0, 0.41)
HPV-11	8.0 (3.6, 12.5)	10.3 (2.9, 17.8)	0.33 (0.00, 0.86)	0.0
HPV-16	19.4 (11.4, 27.5)	18.3 (12.9, 23.7)	0.62 (0.46, 0.78)	0.27 (0.00, 0.54)
HPV-18	9.4 (4.8, 14.0)	11.6 (8.8, 14.4)	0.40 (0.15, 0.65)	0.23 (0.02, 0.48)
HPV-26	11.8 (.,.)	10.0 (5.9, 14.0)	0.00	0.00
HPV-31	20.0 (13.4, 26.6)	14.6 (11.0, 18.1)	0.62 (0.35, 0.89)	0.00
HPV-33	N/A	17.2 (13.0, 21.4)	0.86 (0.61, 1.11)	0.69 (0.32, 1.06)
HPV-35	18.4 (.,.)	18.4 (18.4, 18.4)	100.0	0.00
HPV-39	8.0 (5.8, 10.1)	11.0 (7.0, 14.9)	0.32 (0.03, 0.61)	0.32 (0.03, 0.61)
HPV-40	N/A (all censored)	N/A (all censored)	N/A (all censored)	N/A (all censored)
HPV-42	6.0 (.,.)	12.7 (3.5, 21.9)	0.50 (0.00, 1.19)	0.50 (0.00, 1.19)
HPV-45	8.0 (5.8, 10.1)	11.0 (7.0, 14.9)	0.32 (0.03, 0.61.9)	0.17 (0.00, 0.46)
HPV-51	9.0 (7.7, 10.4)	10.5 (8.4, 12.7)	0.35 (0.14, 0.56)	0.00
HPV-52	13.9 (11.1, 16.8)	14.8 (11.4, 18.3)	0.36 (0.11, 0.61)	0.30 (0.05, 0.56)
HPV-53	16.8 (8.0, 25.7)	13.2 (10.2, 16.1)	0.62 (0.41, 0.83)	0.20 (0.02, 0.42)
HPV-54	8.4 (3.2, 13.6)	10.6 (7.9, 13.2)	0.58 (0.34, 0.82)	0.10 (0.00, 0.28)
HPV-55	5.5 (4.2, 6.7)	7.0 (4.3, 9.6)	0.00	0.00
HPV-56	6.6 (6.0, 7.2)	9.9 (7.0, 12.8)	0.40 (0.13, 0.67)	0.10 (0.00, 0.28)
HPV-57	N/A (no cases)	N/A (no cases)	N/A (no cases)	N/A (no cases)
HPV-58	6.4 (4.9, 7.8)	8.7 (6.8, 10.6)	0.25 (0.14, 0.64)	0.00
HPV-59	7.8 (3.7, 11.9)	10.2 (7.3, 4.3, 9.6)	0.19 (0.00, 0.50)	0.19 (0.00, 0.50)
HPV-66	7.7 (6.4, 9.0)	9.6 (7.2, 12.0)	0.39 (0.12, 0.66)	0.15 (0.00, 0.39)
HPV-68	15.1 (11.4, 18.7)	15.0 (10.9, 19.1)	0.69 (0.32, 1.06)	0.34 (0.00, 0.73)
HPV-73	8.1 (0.7, 15.4)	10.1 (6.5, 13.7)	0.47 (0.10, 0.84)	0.00
HPV-82	9.3 (0.4, 18.2)	12.2 (6.1, 18.2)	0.47 (0.11, 0.84)	0.00
HPV-83	8.5 (1.7, 15.4)	12.0 (6.5, 17.4)	0.5 (0.15, 0.85)	0.19 (0.00, 0.50)
HPV-84	19.4 (11.4, 27.5)	18.3 (12.9, 23.7)	0.23 (0.07, 0.41)	0.12 (0.00, 0.28)

** Estimates from actuarial analysis using the Kaplan-Meier technique*

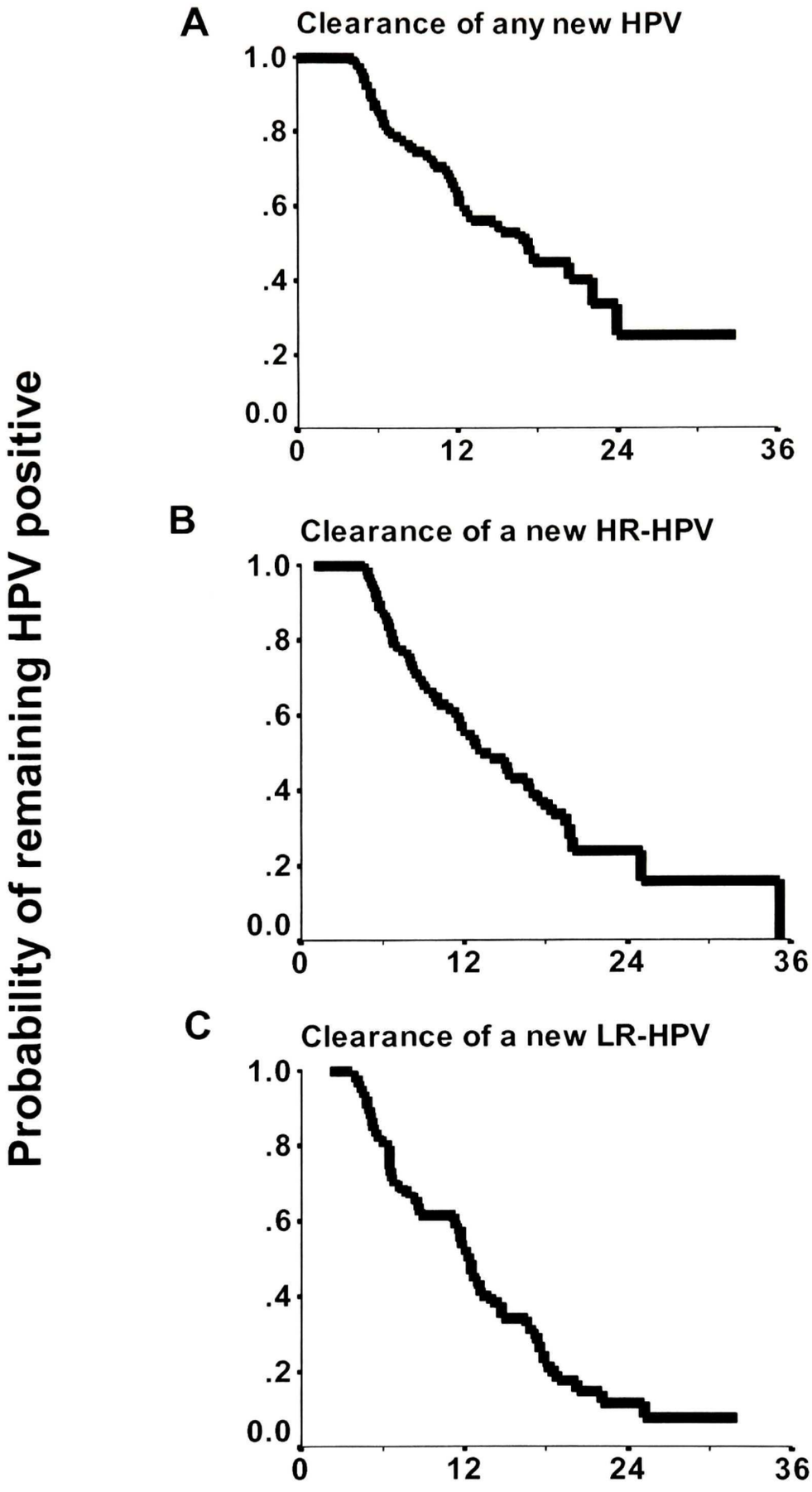
A
Clearance of any new HPV
1.0
.8
.6
.4
.2
0.0
0
12
24
36
B
Clearance of a new HR-HPV
1.0
.8
.6
.4
.2
0.0
0
12
24
36
C
Clearance of a new LR-HPV
1.0
.8
.6
.4
.2
0.0
0
12
24
36
Probability of remaining HPV positive
Time to loss of an incident HPV infection

Figure II.i

Legend for Figure II.i

Probability of remaining HPV positive:
A) Clearance of incident HPV infections (n=157);
B) Clearance of incident type-specific HR-HPV infections (n=123);
C) Clearance of incident type-specific LR-HPV infections (n=73).

Table II.iv

Average duration of type-specific **prevalent** HPV infections with 1- and 2-year probabilities of remaining HPV positive

HPV type	Median retention* time (95%CI) in months	Mean retention* time (95%CI) in months	Proportion* remaining positive at: 1 year (95%CI)	Proportion* remaining positive at: 2 years (95%CI)
HPV-6	8.4 (6.1, 10.8)	7.3 (6.0, 8.6)	0.13 (0.05, 0.30)	0.00
HPV-11	12.2 (6.6, 17.8)	14.7 (7.2, 22.2)	0.60 (0.17, 1.03)	0.00
HPV-16	13.8 (17.0, 20.7)	17.9 (14.4, 21.5)	0.66 (0.5, 0.82)	0.26 (0.08, 0.44)
HPV-18	12.2 (11.0, 13.4)	14.1 (10.4, 17.9)	0.55 (0.31, 0.79)	0.26 (0.02, 0.50)
HPV-26	6.9 (.,.)	10.9 (3.1, 18.8)	0.50 (0.00, 1.21)	0.00
HPV-31	9.9 (1.3, 18.6)	12.9 (7.6, 18.2)	0.46 (0.19, 0.73)	0.23 (0.01, 0.47)
HPV-33	10.5 (7.2, 13.8)	14.6 (8.9, 20.4)	0.50 (0.11, 0.89)	0.33 (0.00, 0.70)
HPV-35	11.4 (.,.)	11.4 (11.4, 11.4)	0.00	0.00
HPV-39	11.7 (2.7, 20.8)	14.8 (7.2, 22.4)	0.50 (0.11, 0.89)	0.33 (0.00, 0.70)
HPV-40	N/A (all censored)	N/A (all censored)	N/A (all censored)	N/A (all censored)
HPV-42	7.3 (.,.)	10.2 (4.5, 15.8)	0.50 (0.00, 1.19)	0.00
HPV-45	11.3 (7.1, 15.4)	11.4 (8.8, 14.0)	0.42 (0.11, 0.73)	0.21 (0.05, 0.47)
HPV-51	7.3 (4.5, 10.1)	10.5 (7.4, 13.5)	0.31 (0.07, 0.55)	0.06 (0.00, 0.18)
HPV-52	13.1 (0.0, 28.8)	16.2 (12.5, 19.9)	0.69 (0.45, 0.93)	0.32 (0.07, 0.57)
HPV-53	14.5 (11.3, 17.7)	15.6 (11.7, 19.5)	0.61 (0.41, 0.81)	0.25 (0.05, 0.45)
HPV-54	13.7 (9.3, 18.1)	17.5 (12.0, 22.9)	0.62 (0.37, 0.87)	0.29 (0.04, 0.54)
HPV-55	7.4 (4.0, 10.8)	6.9 (5.3, 8.5)	0.00	0.00
HPV-56	8.5 (7.1, 9.9)	12.0 (7.4, 16.7)	0.39 (0.10, 0.68)	0.10 (0.00, 0.28)
HPV-57	0.00	0.00	0.00	0.00
HPV-58	12.7 (1.5, 23.9)	16.7 (11.6, 21.9)	0.57 (0.32, 0.82)	0.36 (0.10, 0.61)
HPV-59	4.5 (.,.)	6.5 (3.8, 9.2)	0.50 (0.00, 1.19)	0.50 (0.00, 1.19)
HPV-66	13.5 (2.6, 24.4)	12.7 (7.8, 17.7)	0.57 (0.20, 0.94)	0.14 (0.00, 0.39)
HPV-68	21.0 (.,.)	16.0 (4.6, 27.3)	0.67 (0.14, 1.20)	0.00
HPV-73	7.3 (6.8, 7.8)	6.9 (5.3, 8.4)	0.00	0.00
HPV-82	7.4 (6.4, 8.5)	8.2 (6.1, 10.4)	0.2 (0.05, 0.35)	0.00
HPV-83	7.4 (3.8, 11.1)	6.8 (5.1, 8.5)	0.00	0.00
HPV-84	7.9 (3.5, 12.3)	11.3 (8.2, 14.5)	0.37 (0.15, 0.59)	0.09 (0.00, 0.25)

Table II.v

Average duration of **prevalent** HPV infections, grouped according to oncogenic potential, with 1- and 2-year probabilities of remaining HPV positive

HPV type	Median retention* time (95%CI) in months	Mean retention* time (95%CI) in months	Proportion* remaining positive at: 1 year (95%CI)	Proportion* remaining positive at: 2 years (95%CI)
Any HPV[1] episode	**20.1 (15.7, 24.4)**	**22.0 (19.6, 24.4)**	**0.69 (0.61, 0.77)**	**0.44 (0.36, 0.52)**
HR-HPV[2] episode	**18.2 (13.2, 23.2)**	**19.6 (17.0, 22.3)**	**0.67 (0.59, 0.75)**	**0.36 (0.26, 0.46)**
LR-HPV[3] episode	**14.3 (12.1, 16.5)**	**16.2 (13.9, 18.4)**	**0.59 (0.47, 0.71)**	**0.24 (0.14, 0.34)**

* *Estimates from actuarial analysis using the Kaplan-Meier technique*

[1] *Episode refers to consecutive visits with detection of any HPV type*

[2] *Episode refers to consecutive visits with detection of high-risk HPVs (not necessarily type-specific)*

[3] *Episode refers to consecutive visits with detection of low-risk HPVs (not necessarily type-specific)*

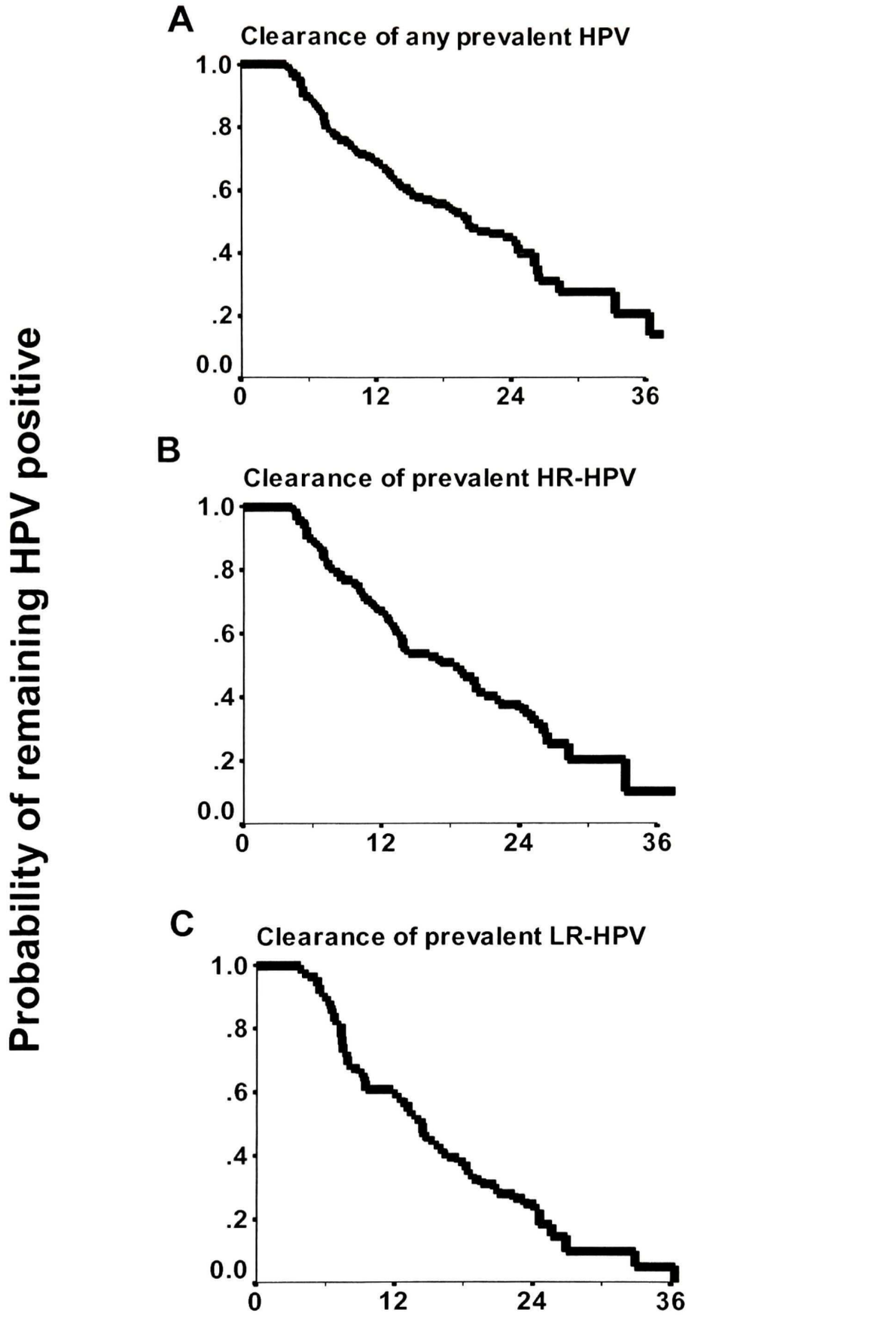
A
Clearance of any prevalent HPV
1.0
.8
.6
.4
.2
0.0
0
12
24
36
B
Clearance of prevalent HR-HPV
1.0
.8
.6
.4
.2
0.0
0
12
24
36
C
Clearance of prevalent LR-HPV
1.0
.8
.6
.4
.2
0.0
0
12
24
36
Probability of remaining HPV positive
Time to loss of a prevalent HPV infection

Figure II.ii

Legend for Figure II.ii

Probability of remaining HPV positive:
A) Clearance of prevalent HPV infections (n=158);
B) Clearance of prevalent HR-HPV infections (n=118);
C) Clearance of prevalent LR-HPV infections (n=80).

APPENDIX III: ADDITIONAL ANALYSES FOR MANUSCRIPT II

Table III.i
Determinants of acquisition of high-risk and low-risk HPV infections:
Model 1: Immune Modifiers

Immune Modifiers	High oncogenic risk Adjusted Hazard Ratio (95%CI)	Low oncogenic risk Adjusted Hazard Ratio (95%CI)
Pack-years smoked		
None	REF	REF
<1 pack-year	0.81 (0.50, 1.32)	1.09 (0.68, 1.75)
1+ pack-years	1.09 (0.68, 1.73)	1.09 (0.67, 1.77)
Alcohol consumption (FUP)		
0 drinks/wk	REF	REF
1-3 drinks/wk	2.22 (1.17, 4.19)	0.83 (0.49, 1.40)
>3 drinks/wk	2.44 (1.24, 4.81)	1.46 (0.86, 2.50)
Duration of OC use		
0 years	REF	REF
<1 year	1.05 (0.60, 1.83)	0.66 (0.39, 1.14)
1-5 years	0.88 (0.54, 1.45)	0.63 (0.39, 1.01)
5+ years	1.01 (0.51, 1.98)	0.53 (0.27, 1.06)
Dairy consumption		
Never	REF	REF
1+ servings/wk	0.25 (0.12, 0.52)	0.82 (0.38, 1.81)

Adjusted for age and sexual activity (number lifetime sexual partners and recent number of new sexual partners); FUP =Time-dependent variables

Table III.ii
Determinants of acquisition of high-risk and low-risk HPV infections
Model 2: Protective Factors

Protective factors	**High- oncogenic risk** Adjusted Hazard Ratio (95%CI)	**Low- oncogenic risk** Adjusted Hazard Ratio (95%CI)
Condom use (FUP)		
Never	REF	REF
Sometimes	0.76 (0.48, 1.21)	0.73 (0.45, 1.18)
Always	0.73 (0.46, 1.14)	0.86 (0.54, 1.37)
Foam use (FUP)		
Never	REF	REF
Sometimes	1.26 (0.58, 2.73)	1.46 (0.56, 3.77)
Always	0.61 (0.15, 2.54)	0.29 (0.04, 2.14)
Wash after sex (FUP)		
Never/rarely	REF	REF
Sometimes	1.08 (0.75, 1.57)	1.03 (0.70, 1.50)
Always	0.54 (0.23, 1.27)	0.80 (0.38, 1.65)

Adjusted for age and sexual activity (number lifetime sexual partners and recent number of new sexual partners); FUP =Time-dependent variables

Table III.iii
Determinants of acquisition of high-risk and low-risk HPV infections
Model 3: Cervical Irritants

Cervical irritants	High-risk Adjusted Hazard Ratio (95%CI)	Low-risk Adjusted Hazard Ratio (95%CI)
Menstrual products		
Pads	REF	REF
Tampons	1.43 (0.80, 2.56)	0.85 (0.47, 1.56)
Pads & tampons	1.10 (0.65, 1.86)	0.77 (0.46, 1.28)
Frequency of sex (FUP)		
Sex <1/wk	REF	REF
Sex 1-2/wk	1.20 (0.75, 1.92)	0.89 (0.57, 1.39)
Sex 3+/wk	1.81 (1.19, 2.75)	0.92 (0.60, 1.42)
History of *Chlamydia* (FUP)		
No	REF	REF
Yes	6.54 (1.46, 29.24)	4.81 (1.05, 22.09)
Vaginal irritation (FUP)		
Never	REF	REF
1-2/since last visit	0.73 (0.49, 1.08)	1.08 (0.73, 1.59)
>2/ since last visit	1.41 (0.84, 2.36)	1.01 (0.56, 1.81)

Adjusted for age and sexual activity (number lifetime sexual partners and recent number of new sexual partners); FUP =Time-dependent variables

APPENDIX IV: ADDITIONAL ANALYSES FOR MANUSCRIPT IV

Table IV.i
Duration of prevalent or incident LSIL according to HPV status at time of diagnosis

HPV status	# events / total # cases	Mean duration in months (95% CI)	Median duration in months (95% CI)
Only LR-HPV	8/9	6.4 (4.9, 7.8)	6.4 (3.6, 9.0)
HR-HPV[1]	12/16	8.5 (5.2, 11.7)	6.2 (5.3, 7.2)
HPV16/18 E-variant[2]	8/12	9.0 (5.7, 12.2)	7.0 (5.1, 8.9)
HPV16/18 NE-variant[3]	2/2	6.5 (0.0, 13.8)	2.8 (-,-)

[1]HR-HPV infections excluding HPV 16 or 18. [2]European variants, and [3]non-European variants for HPV 16 or 18. [4]Calculated with Kaplan-Meier technique (time to SIL clearance).